

# The Development of Theory of Mind and Pragmatics in Adolescents

***Irene Symeonidou***

*Thesis submitted for the degree of Doctor of Philosophy*

*University College London*

*Division of Psychology and Language Sciences*



## Declaration

'I, *Irene Symeonidou* confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.'

Part of the work\* presented in Chapter 2 has been published in the following paper:

Symeonidou, I., Dumontheil, I., Chow, W.-Y., & Breheny, R. (2016). Development of online use of theory of mind during adolescence: An eye-tracking study. *Journal of Experimental Child Psychology*, 149, 81–97. <http://doi.org/10.1016/j.jecp.2015.11.007>

\*The work presented in the thesis differs from the published paper, as I have updated, edited, and extended the chapter where appropriate.



## ACKNOWLEDGMENTS

---

Words cannot express my gratitude to my primary supervisor Dr Richard Breheny. Thank you immensely for your remarkable supervision, your patience, for the priceless knowledge. But most of all thank you for encouraging me to pursue my MSc and ultimately this PhD in the first place. Dr Wing Yee Chow, thank you for your solution-driven guidance and for always having your door open. I could not have asked for a better second supervisor. Dr Iroise Dumontheil, as my third supervisor you went far beyond your 'call of duty'. Thank you for everything you have taught me, for the fantastic opportunity to be part of your lab and for the unique experiences outside of the lab. A special thank you to Dr Heather Ferguson for collaborating with me and making the work in this thesis possible. Thank you for being a mentor and for all your support.

A special thank you to Giulio Dulcinati, Chao Sun, Caterina Paolazzi, Lewis Pollock and Bruno Fernandes for their helpful input and support along the way. Thank you my other PhD colleagues for creating a relaxed, friendly environment to work in and to the iDCNlab members (Annie Brookman-Byrne, Georgina Donati, Lucía Magis-Weinberg, Rosanna Edey and Su Morris) for their invaluable feedback. I would also like to thank Richard Jardine for all his help and support on the admin end of the PhD process and the Linguistics department at UCL for providing an excellent research environment. Thank you to the ESRC for funding my research.

This work would not have been possible without the encouragement and support from my friends and the amazing adventures they created. Thank you to the ones near, Yanna Kyprianou, Christina Varnakidou, Michalis Kapsos, Elina Aletrari, Antonis Theocharides, George Stavrinos, Marina Ergatoude, Raouf Fria, far Simon Loizides, Maria Theodorou, Demetris Charalambous, Harris Constantinou and even further Maria Varkanitsa and Andreas Charidimou. A special thank you to Dimitra Panagopoulos for the daily entertainment and unconditional friendship.

To my family, I thank you from the bottom of my heart for all your unconditional love, support, encouragement, and happy moments you created when I needed them the most. To my parents, my grandparents, my dearest auntie, my uncles, my brothers,

my in-law family, I am forever grateful as I would not have a thesis without you. A special thank you to my mom who has always reminded me that I am capable of anything.

Finally, to my better half and partner in crime Christos, thank you for being a constant source of inspiration and strength. Thank you for always being there for me, for my ups and downs through this roller-coaster ride. I could not have done any of this without you.

## ABSTRACT

---

Theory of Mind (ToM) or 'mentalising' is our human ability to attribute the beliefs, thoughts, desires, intentions and feelings to others, and predict or explain behaviour in terms of these mental states. In the last couple of decades research has shown that brain areas involved in ToM (the 'Social Brain'), undergo changes not only during childhood, but also during adolescence. Numerous studies have provided evidence for structural and functional changes in the Social Brain during childhood and adolescence. Recent findings from behavioural studies suggest a protracted development of ToM through middle childhood and adolescence. However, what factors constrain performance during middle childhood and adolescence are yet to be determined. The current thesis investigates the development of ToM in adolescence and explores what cognitive processes might be developing in parallel to the brain changes that are occurring in the ToM network, through four online ToM tasks.

Chapter 2 examines how children, adolescents and adults apply ToM in real time while performing a variant of the Director task, and the role of inhibitory control in that process. Chapter 3 uses a false-belief task to examine whether adolescents can reliably infer others' (false) beliefs as spontaneously and early as adults. Chapter 4 investigates whether adolescents can use knowledge about a character's basic preferences and higher order desires, even when they are in conflict, to make complex ToM inferences and predict that character's subsequent behaviours as quickly as adults. Chapter 5 examines how adolescents' brain process irony comprehension in real time in comparison to adults' using an ERP paradigm. Additionally, the study explores whether individual differences in empathy are associated with irony processing. Overall, the findings show that ToM is further developing during adolescence and make a theoretical advance as to what specific cognitive processes are still maturing as these brain changes are occurring.

# TABLE OF CONTENTS

---

Acknowledgments .....	5
Abstract .....	7
List of figures .....	11
List of tables.....	12
1 Chapter 1: Introduction .....	13
1.1 Structural development of the Social Brain in adolescence.....	14
1.2 Functional development of the adolescent Social Brain .....	16
1.3 Development of Theory of Mind (ToM) .....	19
1.4 ToM and Executive Function .....	21
1.5 ToM and Empathy .....	22
1.6 The current thesis and summary of experimental chapters .....	23
2 Chapter 2: Adolescents and the Director task.....	26
2.1 Introduction.....	26
2.2 Material and methods .....	31
2.2.1 Participants.....	31
2.2.2 Stimuli and Design .....	31
2.2.3 Procedure .....	34
2.2.4 Data Analysis.....	36
2.3 Results .....	38
2.3.1 Behavioural results .....	38
2.3.2 Eye-tracking results .....	42
2.4 Discussion .....	47
3 Chapter 3: False-beliefs and adolescents.....	53
3.1 Introduction.....	53
3.1.1 The current study .....	55
3.2 Method .....	60
3.2.1 Participants.....	60
3.2.2 Stimuli and design .....	60
3.2.3 Individual measure tasks .....	63
3.2.4 Procedure .....	64
3.2.5 Data analysis.....	65
3.3 Results .....	69
3.3.1 Eye-tracking results .....	69
3.4 Discussion .....	77



4	Chapter 4: Secrets and adolescents.....	83
4.1	Introduction .....	83
4.1.1	The current study.....	84
4.2	Method .....	87
4.2.1	Participants .....	87
4.2.2	Stimuli and design.....	88
4.2.3	Procedure.....	91
4.2.4	Data analysis .....	92
4.3	Results.....	95
4.3.1	Eye-tracking results.....	95
4.3.2	Individual measures results .....	100
4.4	Discussion.....	104
5	Chapter 5: Irony processing and adolescents.....	111
5.1	Introduction .....	111
5.1.1	Theory of mind and adolescents.....	111
5.1.2	Theory of mind and irony processing .....	112
5.1.3	Irony processing in development.....	114
5.1.4	Irony processing and P600.....	115
5.1.5	Empathy and Irony processing.....	116
5.1.6	Current study .....	117
5.2	Method .....	118
5.2.1	Participants .....	118
5.2.2	Stimuli and design.....	119
5.2.3	Questionnaires: empathy and irony comprehension .....	121
5.2.4	Procedure.....	122
5.2.5	Electroencephalogram (EEG) Recording.....	123
5.2.6	Data analysis .....	123
5.3	Results.....	124
5.3.1	ERP results.....	124
5.3.2	Correlations with empathy .....	128
5.4	Discussion.....	129
5.4.1	P600 .....	129
5.4.2	Empathy and irony processing.....	131
5.4.3	Contributions to the Irony processing debate.....	134
5.4.4	Conclusion.....	136
6	Chapter 6: Discussion.....	138

6.1	Summary of findings.....	138
6.2	The big picture.....	140
6.3	Implications for theories on the role perspective information has during language processing.....	142
6.4	Wider implications.....	144
6.4.1	Autism.....	144
6.4.2	Implications for society .....	145
6.4.3	Age of testing.....	147
6.5	Future directions .....	148
6.6	Conclusion .....	150
	References .....	151
	Appendices .....	165
	Appendix I: Chapter 2 supplementary data .....	165
	Experimental items (3-objects) .....	165
	Control items (2-objects).....	168
	Behavioural results .....	171
	Eye-tracking Results.....	172
	Appendix II: Chapter 3 experimental items.....	182
	Appendix III: Chapter 4 experimental items.....	184
	Appendix IV: Chapter 5 experimental items and supplementary data.....	187
	Experimental items for the irony experiment. ....	187
	Supplementary data .....	203

## LIST OF FIGURES

---

<b>Figure 1</b> Instruction screens used to explain the task to participants.....	30
<b>Figure 2</b> Stimuli of the perspective taking task.....	31
<b>Figure 3</b> Percentage error in the Director task for each age group.....	37
<b>Figure 4</b> Percentage error in Go and NoGo trials in the Simple and Complex inhibitory control task for each age group.....	39
<b>Figure 5</b> Target advantage scores per condition over 50ms time bins.....	42
<b>Figure 6</b> Target advantage scores in each age group for incorrect Director Experimental trials. ....	44
<b>Figure 7</b> Visual displays presented to participants in the false-belief task. ....	61
<b>Figure 8</b> The average log(Reality/Belief) score for each condition and age group. ....	67
<b>Figure 9</b> Correlation between participants bias score in the TB and FB condition in the 'container on the' region and each individual measure. ....	75
<b>Figure 10</b> Example stimuli of the secrets task. ....	88
<b>Figure 11</b> The average log(open/secret) score for each condition and age group. ....	96
<b>Figure 12</b> Correlation between participants' age and bias score in the open and the secret conditions.....	98
<b>Figure 13</b> Correlation between participants bias score in the open and secret condition in the ambiguous noun region and scores from each individual measure.....	102
<b>Figure 14</b> Grand average ERPs over twelve electrodes by the critical word in the target sentence for each Story type. ....	125-126
<b>Figure 15</b> Topographic maps of the ERP difference waveform of the Story type (Ironic minus literal) for each age group between 500 and 800 ms from critical word onset. ....	1277

## LIST OF TABLES

---

<b>Table 1</b> Omnibus repeated measures ANOVA of percentage error in the Director task.....	37
<b>Table 2</b> Multiple regression analyses of Director task performance.. .....	40
<b>Table 3</b> Mean target advantage scores for correct trials in each condition for each region.....	43
<b>Table 4</b> Omnibus repeated measures ANOVA of target advantage for correct trials in each region.....	43
<b>Table 5</b> Mean target advantage scores for correct and incorrect Director Experimental trials for each region. ....	45
<b>Table 6</b> Omnibus repeated measures ANOVA F-values of target advantage for correct and incorrect Director Experimental trials in each region. ....	45
<b>Table 7</b> Average time window durations for each condition in the FB task. ....	65
<b>Table 8</b> Estimates and t-values for each time window of interest in the FB task. ....	70
<b>Table 9</b> One sample t-test results for each condition and time window of interest. ....	71
<b>Table 10</b> Means of all individual measure scores and one way ANOVA results .....	73
<b>Table 11</b> Examples of experimental sentences in the secrets task.....	88
<b>Table 12</b> Average time window durations for each condition in the secrets task.....	92
<b>Table 13</b> Estimates and t-values for each time window of interest.....	97
<b>Table 14</b> Means and standard deviations of all individual measure scores and their respective one way ANOVA results.....	100
<b>Table 15</b> Examples stories for each condition and each type of fillers in the irony task. ....	119
<b>Table 16</b> Example stories for irony comprehension questionnaire. ....	121
<b>Table 17</b> Omnibus repeated measures ANOVA .....	128

# 1 CHAPTER 1: INTRODUCTION

---

*...ὁ ἄνθρωπος φύσει πολιτικὸν ζῷον,  
καὶ ὁ ἄπολις διὰ φύσιν καὶ οὐ διὰ τύχην ἥτοι φαῦλός ἐστιν,  
ἢ κρείττων ἢ ἄνθρωπος· ...*  
[Aristotle, Politics 1253a5-9]

“Man is by nature a social animal” (Aristotle, Politics, 1253a5-9). As humans, our everyday lives revolve around social interactions and we seem to suffer in their absence. Understanding others’ behaviour is critical for social interaction. To this end, humans have the ability to attribute the beliefs, thoughts, desires, intentions and feelings to others: what is known as ‘mentalising’ or ‘theory of mind’ (ToM) (Dennett, 1971; 1978; Baron-Cohen, Leslie & Frith, 1985). The progression in imaging techniques (magnetic resonance imaging, MRI) over the last 15 years has allowed significant advances in the understanding of the development of the living human brain. Studies have focused on structural MRI, which examines anatomical changes in the brain (Giedd & Rapoport, 2010; Giedd et al., 1999) and functional MRI (fMRI), which investigates changes in brain activity by detecting changes in blood oxygenation and flow. A substantial number of independent studies have identified specific brain regions that are involved in ToM. The studies used a variety of ToM tasks which involved words, sentences, animations and cartoons and results consistently show activation in a specific network of regions - the ‘Social Brain’- which include the posterior superior temporal sulcus (pSTS), temporo-parietal junction (TPJ), anterior temporal cortex (ACT) and the medial prefrontal cortex (MPFC) (Frith & Frith, 2007; Gallagher & Frith, 2003; Saxe, 2006).

Recently, research has focused on the development of the Social Brain and one of the most prominent finding that has emerged from this research is that the Social Brain does not only undergo changes during childhood, but during adolescence as well (Blakemore, 2008). Adolescence is often defined as the period of life between the start of puberty and the time when self-sufficiency and independence is attained in society (Blakemore & Mills, 2014; Dumontheil, 2015; Kilford, Garrett, & Blakemore, 2016). This

life stage is a period of physical, environmental, behavioural and cognitive change. While it is a time of opportunity it is also the period of time when adult mental disorders such as impulse control, mood, anxiety, and substance-use disorders emerge (Kessler et al., 2005). Moreover, analysis had revealed that most adolescent deaths are as a result of violence, accidents and suicide. Understanding social behavioural development during adolescence may be crucial to better understand the emergence of mental health disorders in certain individuals and behavioural problems which can lead to accidents and violence. Although with advances in MRI and fMRI we now have better understanding of the structural and functional changes in the Social Brain, there has been less focus on behavioural changes in terms of social cognition, in particular ToM, during this period of life.

In this chapter I will first present the findings on the structural development of the Social Brain during adolescence. Secondly, I will summarise the findings from fMRI studies that provide evidence of functional changes in the Social Brain network during adolescence. Thirdly, I will provide an overview of the limited behavioural research investigating ToM during middle childhood and adolescence, our current knowledge of developmental changes in ToM use and the gap in the research that the current thesis addresses. Finally, a summary of the experimental chapters will be introduced.

## **1.1 STRUCTURAL DEVELOPMENT OF THE SOCIAL BRAIN IN ADOLESCENCE**

Early research suggested that structural changes occur in the human brain during adolescence at the cellular level (Huttenlocher, 1979), through a process called synaptic reorganisation. During synaptic reorganisation, synaptogenesis occurs first, followed by synaptic pruning. Synaptogenesis is the process by which new synapses (connections between neurons) are produced in the brain (Blakemore, 2008). In the first months of life, synapses hugely proliferate and by the age of 2 or 3 years old more synapses are present than needed in the brain. Synaptic pruning then occurs whereby synapses that are not being used (depending on the environment that one interacts with) are pruned back and the ones that are being used are strengthened. Huttenlocher (1979) found that in the visual and auditory cortices synaptic pruning begins at around 12 months of age when synaptic density has reached its maximum, and at around 3 and a half years in the

prefrontal cortex (PFC). Crucially, whilst synaptic pruning seems to end by the age of 12 in the auditory cortex, in the PFC it continues during adolescence and beyond (Huttenlocher, 1979). A more recent study by Petanjek and colleagues (2011) found evidence that synaptic elimination occurs not only throughout adolescence but also until the mid-30s. Therefore, synaptic pruning in the PFC, an area that is part of the Social Brain, occurs throughout childhood, adolescence and beyond.

White and grey matter in the PFC undergo changes throughout adolescence as well. MRI studies consistently suggest that white matter volume increases linearly during childhood and adolescence in numerous brain regions including the PFC, frontal and parietal cortices (Aubert-Broche et al., 2013; Barnea-Goraly et al., 2005; Giedd et al., 1999; Sowell, Thompson, Holmes, Jernigan, & Toga, 1999). It has been suggested that the white matter increase could indicate continued axonal myelination throughout childhood and adolescence (Blakemore, 2008; Yakovlev & Lecours, 1967). A longitudinal study that scanned 103 individuals aged 5-32 years old at least twice (Lebel & Beaulieu, 2011) found an increase in white matter volume between the ages of 5 and 25.

Unlike white matter, evidence suggests that grey matter in the prefrontal cortex increases during childhood and then decreases during adolescence (Aubert-Broche et al., 2013; Tamnes et al., 2013). Tamnes, Walhovd, Dale and colleagues (2013) scanned 85 participants twice and found that grey matter volume decreased between 8 and 22 years of age. Similarly, in Aubert-Brioche and colleagues (2013) grey matter volume decreased between 10 and 20 years of age. Furthermore, it has been shown that cortical thickness matures earlier in sensory and motor regions while frontal and temporal lobes mature at a later stage (Shaw et al., 2008). The MPFC specifically seems to be one of the last regions to reach its peak in cortical thickness (Shaw et al., 2008). Various interpretations for the developmental changes in grey matter have been proposed. One explanation is that the grey matter volume decrease during adolescence, evident in MRI studies, is due to intracortical myelination (Burnett, Sebastian, Cohen Kadosh, & Blakemore, 2011; Paus, Keshavan, & Giedd, 2008). Another interpretation is that the grey matter changes are reflective of the synaptic reorganisation. More specifically, the idea is that the increase in volume implicates synaptogenesis and the decrease that follows in adolescence reflects synaptic pruning during which synapses are eliminated

(Blakemore, 2008). However, it is debated whether the above changes can be captured as volumetric changes in MRI scans (Paus et al., 2008).

A recent longitudinal study (Mills, Clasen, Giedd, & Blakemore, 2014) investigated the structural developmental trajectory of the brain areas specifically involved in ToM (ATC, MPFC, TPJ, pSTS). The study examined brain scans of individuals between the ages of 7 and 30 years old using three structural measures: grey matter volume, cortical thickness and surface area. Results showed that while grey matter volume decreased in MPFC, TPJ and pSTS from childhood into the early twenties, grey matter volume in the ATC increased until around the age of 12 and then decreased. Cortical thickness decreased in MPFC, TPJ, and pSTS until early twenties but increased in the ATC until early adulthood. For all four brain areas surface area increased until late childhood or early adolescence and decreased thereafter into the early twenties (Mills et al., 2014). Overall, the evidence suggest that brain areas implicated in social cognition processes, and specifically in ToM, undergo several structural changes during childhood and adolescence.

## **1.2 FUNCTIONAL DEVELOPMENT OF THE ADOLESCENT SOCIAL BRAIN**

In addition to the structural changes, numerous fMRI studies have shown changes in functional recruitment of the ToM network between adolescence and adulthood. Using a variety of ToM tasks, studies have consistently found that in comparison to a control task, MPFC activity decreases between adolescence and adulthood (see Blakemore, 2008, 2012). For example, when judging sincerity vs. irony through cartoons, children and adolescents activated the MPFC, left IFG and the right pSTS more than adults (Wang, Lee, Sigman, & Dapretto, 2006). These results were replicated in another study where participants had to judge scenarios that involved either physical causality or intentional causality (Blakemore, den Ouden, Choudhury, & Frith, 2007). During the intentional causality condition, when participants had to think about their own intentions, adolescents showed greater activation in the MPFC than adults whereas adults activated the right STS more than adolescents. Similarly, Pfeifer and colleagues (2007) investigated children (9.5-10.8) and adults (23-31.7) when judging whether phrases described themselves or Harry Potter. In the self-knowledge retrieval



condition, children showed greater activation in the MPFC and ACC than adults (Pfeifer, Lieberman, & Dapretto, 2007).

In a study that involved judging emotional response during social vs. basic emotion scenarios Burnett and Blakemore (2009) found that activity in the MPFC decreased with age during the social emotion condition and activity in the left anterior temporal cortex (ATC) increased with age. Burnett and Blakemore (2009) conducted the same experiment and demonstrated that adolescents have a stronger functional connectivity between MPFC and pSTS/TPJ than adults (Burnett & Blakemore, 2009). Moreover, Goddings and colleagues (2012) used to the same paradigm to investigate whether pubertal hormones have an effect on social emotion processing. While age was correlated with lower MPFC activity during social emotion processing independently of age, results also revealed that increased hormone levels correlated with higher left ATC activity independent of age, suggesting that age and pubertal hormones have different effects on the adolescent mentalising network (Goddings, Heyes, Bird, Viner, & Blakemore, 2012).

It has been argued that the range of ToM tasks that have been used, e.g. thinking about social emotions such as embarrassment (Burnett & Blakemore, 2009), thinking about one's own intentions to carry out an action (Blakemore et al., 2007), attributing emotional states of others (Gunther Moor et al., 2011), evaluating affective mental states of actors in video clips (Vetter, Altgassen, Phillips, Mahy, & Kliegel, 2013) and responding to cartoon scenarios that involve affective ToM (Sebastian et al., 2012), all involve an explicit judgment or response about the mental states of another or about one's own mental state (Dumontheil, 2015; Dumontheil, Hillebrandt, Apperly, & Blakemore, 2012). In an fMRI study, Dumontheil and colleagues (2012) tested 11-16 year old adolescents and adults using an adapted version of Director task (DT) paradigm in order to investigate neural substrates of ToM use during an implicit task. During the task participants were presented with a set of shelves containing some objects. A director standing on the other side of the shelves instructed participants to move objects on the shelf. Some of the objects were blocked off from the director's point of view, therefore participants had to take into consideration the director's ignorance of these objects when following the director's instructions. In addition, unlike the previous variants of

the DT (Dumontheil, Apperly, & Blakemore, 2010; Keysar, Barr, Balin, & Brauner, 2000), the current task included a second director who was standing on the same side as the participant. In other words, participants had to also take into account whether the director was present on the same side of the shelves as them or on the other side. In a control condition participants performed the same task in the absence of the director and were told the instructions would only refer to items on slots without a back panel, controlling for general cognitive demands of the task. Results showed that compared to the control No Director condition, the Social Brain regions, specifically the MPFC and along the superior temporal sulcus, showed greater activation in the Director than the control condition. Moreover, results revealed a difference in activation of the MPFC between the adolescent and adult groups. Adults showed greater activation in the MPFC only during the experimental trials where participants had to choose one out of three objects, one of which was occluded from the Director and was the best linguistic fit for the instruction; in comparison to the control trials where they had to select one of two objects that were both visible to the director and only one object fit the instruction (Dumontheil, Küster, Apperly, & Blakemore, 2010). Adolescents, however, showed MPFC activation in both experimental and control trials of the Director condition (Dumontheil et al., 2012). The authors propose that adults show more specificity of MPFC activation for ToM use and that adolescents might be showing “over-mentalising” activation in ToM tasks.

In sum, evidence from fMRI studies suggest changes in activity in the ToM network during adolescence. In explicit ToM tasks neuroimaging research has consistently shown that compared to a control condition, MPFC activity was greater in adolescents than the adults during the condition that involved ToM processes. Moreover, activity in the ATC increases with age or with levels of pubertal hormones indicating a more advanced stage of puberty. Added to that, evidence suggests that during implicit tasks adults show an increased specificity of MPFC activation for ToM in comparison to adolescents.

### **1.3 DEVELOPMENT OF THEORY OF MIND (ToM)**

Most of the developmental research on ToM has focused on the preschool years and has shown that children's ability to attribute false-beliefs to others emerges at about the age of 4 and 5 years old (Baron-Cohen, Leslie, Frith, 1985; Happé, 1995, Perner, Leekam, & Wimmer, 1987; Wimmer & Perner, 1983). Although children can pass standard 'explicit' false belief tasks from the age of 4, younger children shown an appreciation of false belief from as young as 13 months. Evidence has shown that infants have a sensitivity to other people's false-beliefs as they observe the false-belief location longer as opposed to the location where a certain item is currently located (Onishi & Baillargeon, 2005 in 15 month olds; Surian et al., 2007 in 13 month olds). This suggests children have an expectation that the agent will reach in the location where he/she falsely believes the object is (Baillargeon, Scott, & He, 2010).

Since research has demonstrated that ToM skills are maturing between the ages of 2 and 7 years old (Wellman, Cross, & Watson, 2001), it is unsurprising that little empirical research has focused on ToM development beyond childhood. Moreover, the ToM paradigms that have been used to test pre-school children are not suitable to test older children and adolescents as they elicit ceiling performance (Dumontheil, 2015; Dumontheil, Apperly, et al., 2010). However, given the structural changes occurring in the brain regions involved in ToM during the adolescent years the question arises of whether there is an impact at the cognitive level on the development of mental state understanding beyond the pre-school years. A few recent findings indeed suggest that the online use of ToM shows a prolonged development through middle childhood (Devine & Hughes, 2013; Epley, Morewedge, & Keysar, 2004; Surtees & Apperly, 2012) and adolescence (Dumontheil, Apperly, et al., 2010; Vetter, Leipold, et al., 2013; Vetter, Altgassen, et al., 2013).

Devine & Hughes (2013) investigated ToM development beyond the pre-school years in children between 8-13 years of age using a novel ToM reasoning task – the Silent Film task. The task is a film-based analogue to the Strange stories task (Happé, 1994) that is well suited to exploring the ability to understand others' beliefs in children beyond the age of 7. The participants were presented with short film clips that depicted scenarios involving deception, false belief, and misunderstanding and were asked to

explain the behaviour of a character. After controlling for verbal ability and socioeconomic status, performance on the task significantly improved with age, suggesting that ToM abilities do continue to develop through middle childhood.

Using the Story Comprehension test, (Channon & Crawford, 2000), Vetter and colleagues (2013) tested ToM in adolescents (12-15) and young adults (18-22). The story types included situations of threat, dare, white lies, irony and pretence. Results showed age group differences, in particular, the adolescent group performed significantly worse in the white lie, dare and elaborated irony stories, which the authors suggest are the stories which involved more advanced ToM processes (Vetter, Leipold, et al., 2013). No age-related differences were found in the stories assessing pretence, threat and excuse, which, according to the authors, require basic ToM abilities. This study provides evidence of ToM development in mid-adolescence, however it is unclear whether there may be further changes in late adolescence as there was no adult (24+ years old) group.

In the tasks used in both Devine & Hughes (2013) and Vetter and colleagues (2013) participants had to make explicit judgements about the mental states of a character in a certain story. In contrast, Dumontheil and colleagues (2010) tested 7 to 27 year-old female participants using a more implicit paradigm. Moreover, they tested two child groups, two adolescent groups and one adult group allowing for a clearer developmental picture between children, both young and old adolescents and adults. The study used a computerised version of Keysar and colleagues' (2000) Director task where participants were presented with a 4 x 4 grid that contained various objects in different slots. A director who was standing on the other side of the shelves, gave them instructions on which objects to move and where to move them. While some of the shelves had no background and were visible to the director, other objects were on shelves that had a back panel, blocking them from the director's view. Therefore, participants had to take the director's ignorance of these objects into consideration when following his instructions. Participants also performed a control No Director condition to control for general cognitive demands of the task. In the No Director condition the task was exactly the same with the exception of the director's absence. Participants were told they would hear instructions which would only refer to items on slots without a back panel. The results showed that in the director-present condition,

but not the control No Director condition, the late adolescent group (14-17) made more errors than adults, suggesting that ToM use differs between adolescents and adults when certain executive demands are controlled for.

While there is a well-established consensus of how ToM is defined, a recent model proposes that ToM is multidimensional process that is comprised of two main components: i) cognitive ToM which refers to the ability of inferring others beliefs, and ii) affective ToM which refers to the ability of inferring others emotions (Simone G. Shamay-Tsoory, Harari, Aharon-Peretz, & Levkovitz, 2010). Interestingly, there is evidence to suggest that adolescents' performance is poorer on affective ToM tasks in comparison to cognitive ToM tasks (Sebastian et al., 2012; Vetter, Altgassen, et al., 2013). Sebastian and colleagues (2012) observed that adolescents (11-16) made significantly more errors than adults (24-40) in cartoon scenarios that involved affective ToM where participants had to infer how one story character would react to the other character's affective state. Conversely, there were no age-group differences in cartoons that involved cognitive ToM where participants had to infer how one character would react to the other based on beliefs and intentions.

#### **1.4 ToM AND EXECUTIVE FUNCTION**

As previous findings from both behavioural and brain data suggests associations between executive function and ToM, the current thesis also aims to establish to what extent ongoing development of executive function processes play a role in ToM development in adolescents. Executive function (EF) refers to the processes that allow us to control and coordinate behaviour (Luria, 1966; Shallice, 1982) such as inhibition and working memory. Neuropsychological research has demonstrated that EF has a protracted period of development, which begins in early childhood (~2 years old) and continues into young adulthood, and each sub-component of EF develops at its own rate (Diamond, 2002). Behavioural studies investigating executive function in adolescence found that it is still under development until mid to late adolescence (Luna, Padmanabhan, & Hearn, 2011). This has been studied using tests of inhibitory control (Leon-Carrion, García-Orza, & Pérez-Santamaría, 2004), working memory (e.g. (Conklin, Luciana, Hooper, & Yarger, 2007; Gathercole, Pickering, Ambridge, & Wearing, 2004;

Luna, Garver, Urban, Lazar, & Sweeney, 2004b)), and decision making (Luciana, Conklin, Hooper, & Yarger, 2005).

Previous research has shown an association between inhibitory control (IC) and the application of ToM in children (3-4 years old in Carlson, Moses, & Claxton, 2004; meta-analysis of 3-6 years old in Devine & Hughes, 2014; 4-10 years old in Lagattuta, Sayfan, & Harvey, 2014; 4-9 years old in Lagattuta, Sayfan, & Blattman, 2010; 2-5 years old in Nilsen & Graham, 2009) and in adults (Brown-Schmidt, 2009b). One study investigated the relation of IC and ToM in adolescence (Vetter, Altgassen, et al., 2013) and found that IC, as measured with an anti-saccade task, predicted age-related variance in affective ToM. Studies have also shown that working memory (WM) has an effect on ToM processing. Specifically, studies have found that lower WM capacity or increased WM load results in poorer performance on ToM tasks, such as perspective taking (Bull, Phillips, & Conway, 2008; Cane, Ferguson, & Apperly, 2017; Lin, Keysar, & Epley, 2010; McKinnon & Moscovitch, 2007; Nilsen & Graham, 2009; Schneider, Lam, Bayliss, & Dux, 2012).

Associations between executive function and ToM will be explored in Chapters 2, 3 and 4. Specifically, Chapter 2 will investigate the role of inhibitory control during perspective taking. Chapters 3 and 4 will explore whether individual measures in IC and WM are associated with ToM in adolescence.

## **1.5 TOM AND EMPATHY**

In addition to EF, empathy has also been shown to be a factor in ToM abilities. Empathy is the ability to share the emotional states of others (Bernhardt & Singer, 2012; Singer, Critchley, & Preuschoff, 2009). Evidence from neuroimaging has suggested that ToM processes and empathy processes activate overlapping neuronal networks such as the MPFC, TPJ and temporal poles (Völlm et al., 2006). Moreover, in an ERP study (Ferguson, Cane, Douchkov, & Wright, 2015) results revealed that participants with high levels of empathy are better at making inferences according to someone else's false-belief with a corresponding modulation of the ERPs, whereas people with lower levels of empathy were more likely to interpret events egocentrically (Ferguson, Cane, et al., 2015). Additionally, behavioural data suggests that empathy is still developing during

adolescence (Sebastian et al., 2012). Chapters 3 and 4 will investigate whether individual measures in empathy are associated with ToM in adolescence. In Chapter 5 I will investigate whether individual measures in empathy are associated with irony.

## **1.6 THE CURRENT THESIS AND SUMMARY OF EXPERIMENTAL CHAPTERS**

Although evidence from previous studies suggests a protracted development of ToM, the factors that constrain performance during middle childhood and adolescence are yet to be determined. Given that children as young as 4 or even younger can pass false belief tests, it is clear that adolescents have the conceptual ability to think about the actions of another agent in terms of underlying beliefs, desires and intentions. The key question is what is occurring at the cognitive level in parallel with these structural and functional changes in the brain? One possible explanation for the differences in MPFC activity discussed above that has been proposed is that adolescents and adults use a different cognitive strategy (Blakemore, 2008; Dumontheil, 2015). Specifically, one possibility is that adults retrieve ‘social scripts’ in the ATC based on previous experiences, while adolescents who have less experience process new computations in MPFC to interpret that same social situation (Dumontheil, 2015).

In addition to the above suggestion, there is some evidence that EF abilities are a possible factor (Vetter, Altgassen, et al., 2013). This is perhaps not surprising since, as stated above, EF abilities are also changing through adolescence. Moreover, there is also an indication that while perhaps adolescents’ cognitive ToM is intact, their affective ToM might still be maturing (Sebastian et al., 2012). Beyond these possible proposals however, there has been no further theoretical advance on what cognitive (psychological) processes might still be developing in parallel to the changes that are occurring in the Social Brain.

In the current thesis, I address these issues with a series of ToM tasks that involve on-line measures as opposed to the off-line measures used in the studies mentioned above. Although ToM paradigms have been developed that involve more sensitive on-line measures such as reaction times (Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006; Back & Apperly, 2010), brain responses (Ferguson, Cane, et al., 2015; Geangu, Gibson, Kaduk, & Reid, 2013; Meinhardt, Kühn-Popp, Sommer, & Sodian, 2012) and eye-

tracking (Epley et al., 2004; Ferguson, Apperly, Ahmad, Bindemann, & Cane, 2015; Ferguson & Breheny, 2012; Hanna, Tanenhaus, & Trueswell, 2003; D. Heller, Grodner, & Tanenhaus, 2008; Keysar et al., 2000), studies have been limited to either adults (Keysar, Lin, & Barr, 2003; Lin et al., 2010; Wu & Keysar, 2007), or younger children (5-6 years old in Nadig & Sedivy, 2002; 6-10 years old in Surtees & Apperly, 2012; 4-12 years old in Epley, Morewedge, & Keysar, 2004). These time-sensitive methods have not been used to study ToM in adolescents.

In **Chapters 2-4** eye-tracking is used in different visual world paradigms to measure the time-course of ToM processes in different types of ToM tasks. It is widely accepted that eye-fixations can indicate how we process information during spoken language comprehension (Altmann & Kamide, 1999; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). By measuring when participants fixate their gaze on an object we can identify which object they are considering as a possible referent at a given point in time. **Chapter 5** on the other hand, will use event-related brain potentials (ERPs), which are also time-sensitive measurements that provide continuous information on cognitive processes underlying a response.

In sum, the thesis will investigate the development of ToM in adolescence and what factors account for this age-related improvement. Additionally, in this thesis I aim to make a theoretical advance as to what cognitive processes are changing as these brain changes are occurring during this period of life. While this introduction gives a thorough description of the structural and functional changes that are occurring in the Social Brain during adolescence, the current thesis will not aim to investigate a direct relationship between the cognitive processes and the changes occurring in the brain. The review of the literature is presented in order to show the robust evidence that these changes are occurring during adolescence - which is quite a recent finding in the field - and why there is a strong motivation to investigate whether there are differences at the cognitive level beyond the childhood years. **Chapter 2** examines and compares how children, adolescents and adults apply ToM in real time while performing a variant of the Director task (Dumontheil et al., 2010; Keysar et al., 2000). Moreover, the role of IC in perspective taking is investigated. **Chapter 3** uses a false-belief task to examine whether adolescents can reliably infer others' (false) beliefs as spontaneously and early as adults, even



without being given an explicit instruction to do so. **Chapter 4** investigates whether adolescents can use knowledge about a character's basic preferences and higher order desires, even when they are in conflict, to make complex ToM inferences and predict that character's subsequent behaviours as quickly as adults. In both **Chapters 3 and 4** individual differences in IC, WM and empathy are measured to explore whether they are factors in constraining performance during the two different ToM tasks. **Chapter 5** examines how adolescents' brain process irony comprehension in real time in comparison to adults' using an ERP paradigm. Additionally, the study explores whether individual differences in empathy are associated with irony processing. Overall, the findings in the chapters show that ToM is further developing during the adolescent years and crucially, point to specific abilities that might still be maturing.

## 2 CHAPTER 2: ADOLESCENTS AND THE DIRECTOR TASK

---

### 2.1 INTRODUCTION

In Chapter 1 I reviewed the findings of cognitive neuroscience research of the last couple of decades and how it has shown that brain areas involved in Theory of Mind (the ‘Social Brain’), undergo changes not only during childhood, but also during adolescence (Burnett et al., 2011). Specifically, I gave an overview of the studies that provided evidence for structural and functional changes in the Social Brain during childhood and adolescence. Moreover, I presented the few studies that have found evidence that ToM is applied more robustly or accurately with age through middle childhood (Devine & Hughes, 2013; Epley et al., 2004; Serena Lecce, Bianco, Devine, Hughes, & Banerjee, 2014; Surtees & Apperly, 2012; J. J. Wang, Ali, Frisson, & Apperly, 2015) and adolescence (Dumontheil, Apperly, et al., 2010; Vetter, Altgassen, et al., 2013). Although previous studies have shown that children’s performance on paradigms such as false-belief tasks reaches ceiling at around the age of 5 (Surian, Caldi, & Sperber, 2007; Wellman, Cross, & Watson, 2001) this recent evidence of further development raises the question of what factors affect older children and adolescents’ successful application of such abilities.

The Director paradigm has been used to investigate the ability to take the perspective of another individual into account in a communicative context (Apperly, Back, Samson, & France, 2008; Brown-Schmidt & Hanna, 2011; Fett et al., 2014; Keysar et al., 2000, 2003). In these studies, the participant interacts with another agent (a ‘director’) to act on a set of objects (Director paradigm; **Figure 1**). Crucially, some of the objects are blocked off from the director’s point of view and are visible only to the participant. Thus, when the director talks about an object (e.g. ‘the large ball’, **Figure 2**), the participant should ignore any object that is not visible to the director and instead select a referent from what is in ‘common ground’, i.e. what is visible to both the participant and the director. This paradigm requires the participant to infer the speaker’s referential intention (a mental state) based on beliefs that differ from their own due to the speaker’s ignorance of the presence of an object that would be a

potential referent for the instruction given. For example, in the set-up shown in **Figure 2**, the participant, but not the director, sees a third ball which best fits the description 'the large ball' (the basketball) and has to discount it as the intended referent as the director does not know about that ball.

The Director paradigm is useful for the study of the development of the social brain as it can be used to measure the application of aspects of Theory of Mind (ToM) without asking participants to make an explicit judgement about their own or someone else's perspective or thoughts. Given that even adults manifest less-than-ceiling performance in the Director paradigm (Keysar et al., 2000; Brown-Schmidt & Hanna 2011), it is well suited for exploring how ToM abilities develop across childhood and adolescence.

Early visual-world eye-tracking studies using the Director paradigm demonstrated that adult participants are not able to ignore objects that they have privileged visual access to and that they are less accurate in choosing the object that is visually available to both the participant and the director compared to control conditions (Epley et al., 2004; Keysar et al., 2000, 2003). Keysar and colleagues explained these results by suggesting that participants have an initial bias to take an egocentric perspective and that their initial interpretation of the instruction is later adjusted according to the speaker's knowledge by a second process, which corrects any manifest errors.

In contrast, studies using very similar paradigms have found evidence that participants are able to integrate information about the speaker's beliefs immediately (Hanna et al., 2003; Heller et al., 2008). This is comparable with findings that individuals may rapidly and automatically compute what other people see (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010). It seems clear that adults are able to integrate information to some extent about what their interlocutor knows from the earliest stages of referential language processing, even though they are liable to attend to objects that are not known to the speaker (Brown-Schmidt & Hanna, 2011). Among adults, performance on perspective-taking tasks has been shown to be affected by factors such as the nature of the verbal stimulus (Barr, 2008), the extent to which the interaction offers information about the interlocutor's mental state (Brown-Schmidt,

2009a), cultural background (Wu & Keysar, 2007), inhibitory control (Brown-Schmidt, 2009b; Nilsen & Graham, 2009), and mood (Converse, Lin, Keysar, & Epley, 2008).

As reviewed in Chapter 1, Dumontheil, Apperly, & Blakemore (2010) tested 7 to 27 year-old female participants' performance using a computerised version of Keysar and colleagues' (2000) Director task (DT). Participants were presented with a 4 x 4 grid that contained various objects in different slots. A director (an avatar) gave participants instructions about which objects to move and where to move them. As in the original design, some of the slots on the shelves were occluded so the director could not see what was present on those shelves. The study also included a No-Director condition where participants performed the same task in the absence of a director. In this condition they were told that the instructions would not refer to items on slots with a grey background. According to Dumontheil and colleagues (2010), the difference between the two conditions is that while only executive function processes are required for the participants to perform well on the task in the No-Director condition, both ToM processes and executive function processes are required in the Director condition because participants have to take someone else's perspective into account in order to perform well.

The authors found that error rates were higher in the late adolescent group (14-17 year olds) than in the adult group in the Director condition, but they did not differ between the late adolescent and adult groups in the No-Director condition, which relies on executive functions only. This suggests that the late adolescent group differs from adults in their application of Theory-of-Mind inference when certain executive function demands are controlled for.

In contrast to the error rates, the authors reported that reaction time on the critical trials in the Director condition remained the same across all age groups. Further, they found that reaction times were longer in the No-Director condition than in the Director condition. They comment that this difference suggests that participants approach the Director task in a way that is more efficient than just applying an arbitrary rule. They propose that the difference in accuracy between the late adolescent and adult groups may not stem from how efficiently or rapidly they can make referential inferences but from their propensity to integrate information about the speaker's

beliefs in making those inferences. These suggestions point to a response to the question raised above as to what factors are responsible for the better application of Theory of Mind abilities in this stage of development. The present study uses visual-world eye-tracking measures on a variant of the Director task employed in Dumontheil and colleagues (2010).

Although eye-tracking has previously been used in variants of the Director task, samples of these studies have been limited to either adults (Keysar et al., 2003; Lin et al., 2010; Wu & Keysar, 2007), or younger children (5-6 years old in Nadig & Sedivy, 2002; 4-12 years old in Epley, Morewedge, & Keysar, 2004). The present study affords an opportunity to examine and compare how children, adolescents and adults apply ToM in real time through monitoring their eye-gaze while they are performing the task. Specifically, the aim is to explore the time-course of the decision process that yields both correct and incorrect responses in participants.

Further, the present study also aimed to understand whether ongoing development of executive function processes, specifically inhibition, plays a role in perspective taking. As highlighted in Chapter 1 previous studies have found a correlation between inhibition and the application of ToM inference in children (3-4 years old in Carlson, Moses, & Claxton, 2004; 4-10 years old in Hansen Lagattuta, Sayfan, & Harvey, 2014; 4-9 years old in Lagattuta, Sayfan, & Blattman, 2010; 2-5 years old in Nilsen & Graham, 2009; meta-analysis of 3-6 years old in Devine & Hughes, 2014) and in adults (Brown-Schmidt, 2009b). A study by Vetter, Altgassen, Phillips, Mahy and Kliegel, (2013) has found that inhibitory control, as measured with an anti-saccade task, predicted age-related variance in affective ToM in adolescence. No other study, however, has looked at the role of inhibitory control in adolescents' ability to inhibit one's perspective and consider someone else's. Therefore, we measured participants' inhibitory control through two variants of the Go-NoGo task in order to investigate whether individual differences in inhibitory control can account for their performance in the Director task.

Based on the evidence of ongoing functional development of neural processes related to ToM and the results of Dumontheil et al. (2010), we expected that age-related differences in participants' accuracy would be greater in the Director condition than in the No-Director condition. We also expected to see age-related differences in inhibitory

control, which may account for participants' performance in the Director task to some extent. We measured the extent to which participants considered the target object relative to the distractor by computing target advantage scores, which is the average probability of looks at the target object minus the average probability of looks to the distractor (Brown-Schmidt, 2009b; Kronmüller & Barr, 2015). This allowed us to examine and compare how each age group used ToM to identify the referent.

According to the Perspective-Adjustment model proposed by Keysar and colleagues, participants have an initial egocentric bias, which is then adjusted, during a second process, to the speaker's perspective. Therefore, this theory predicts that participants will fixate on the distractor first and, if the second process of adjusting is successful (i.e. on correct trials), participants will then fixate their gaze on the target object and give a correct response. However, if the second process fails and perspective taking does not occur then we would expect participant's gaze to remain on the distractor, not to fixate on the target, and eventually give an incorrect response. As such, it is predicted that while both adolescents and adults would fixate on the distractor object initially, demonstrating an initial egocentric bias, adults would then be more likely to switch to fixating their gaze on the target object and do so more quickly than adolescents. In other words, both adolescents and adults would have a smaller target advantage initially, but adults would have a bigger target advantage score than adolescents in earlier time regions.

Other models of perspective-taking found in Hanna and colleagues (2003) and Brown-Schmidt and colleagues (2008) suggest that perspective information is available from the outset but that the extent to which it is integrated on-line depends on how sensitive a participant is to such information (Brown-Schmidt, 2012) or their ability to integrate that information with linguistic and visual contextual information (Hanna et al., 2003). According to these proposals, an incorrect response may reflect a simple failure to integrate perspective from the start of a trial. If this is the case, patterns of eye gaze may differ from the onset of the critical phase of the instruction between correct and incorrect trials.

## 2.2 MATERIAL AND METHODS

### 2.2.1 Participants

Sixty-five participants took part in this study, constituting three age groups: Younger adolescent ( $n=14$ , 9-13 years old,  $M = 11.2$ ,  $SD = 1.2$ ), Older Adolescent ( $n=28$ , 14-17.9 years old,  $M = 16.2$ ,  $SD = 1.1$ ) and Adult groups ( $n=23$ , 19-29 years old,  $M = 23.0$ ,  $SD = 2.8$ ). The age groups were divided to match the age groups of Dumontheil et al. (2010). Adult participants were recruited from the UCL Psychology participant pool while younger and older adolescents were recruited from London schools. All participants were native English speakers. Data from three adults were excluded due to technical errors. We also excluded data from two adolescents who did not understand the instructions of the No-Director condition. Verbal ability was measured in all participants using the vocabulary subtest of the Weschler Abbreviated Scale of Intelligence (WASI, Wechsler, 1999). Average verbal IQ scores were 124 ( $SD = 8.0$ ) for adults, 118 ( $SD = 6.9$ ) for older adolescents and 123 ( $SD = 10.2$ ) for younger adolescents. A one way ANOVA showed a significant difference in verbal IQ scores between groups ( $F(2,57) = 5.97$ ,  $p = .03$ ). Bonferroni corrected post-hoc comparisons showed that older adolescents performed worse than adults ( $p = .04$ ) but verbal IQ did not significantly differ between younger adolescents and older adolescents ( $p = .22$ ) or young adolescents and adults ( $p = 1$ ).

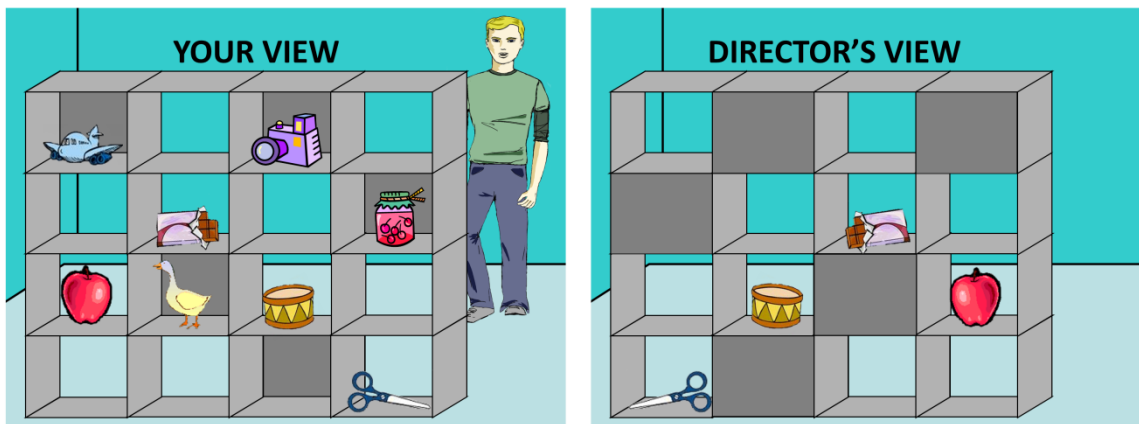
Parents/guardians of all adolescent participants as well as the adult participants were given information sheets prior to the study. Informed consent was obtained from the parent/guardian for all adolescent participants and from all adult participants. This study was approved by the UCL Research Ethics Committee.

### 2.2.2 Stimuli and Design

#### 2.2.2.1 Director task

The present experiment had two within-participant factors: Condition (Director, No-Director) and Trial type (Control, Experimental) and one between-participants factor: Age group (Young Adolescent, Older Adolescent, Adult). The task design and stimuli were based on a previous study by Dumontheil, Apperly and Blakemore (2010; see also Apperly et al., 2010). Participants were presented with a visual scene with a 4 x

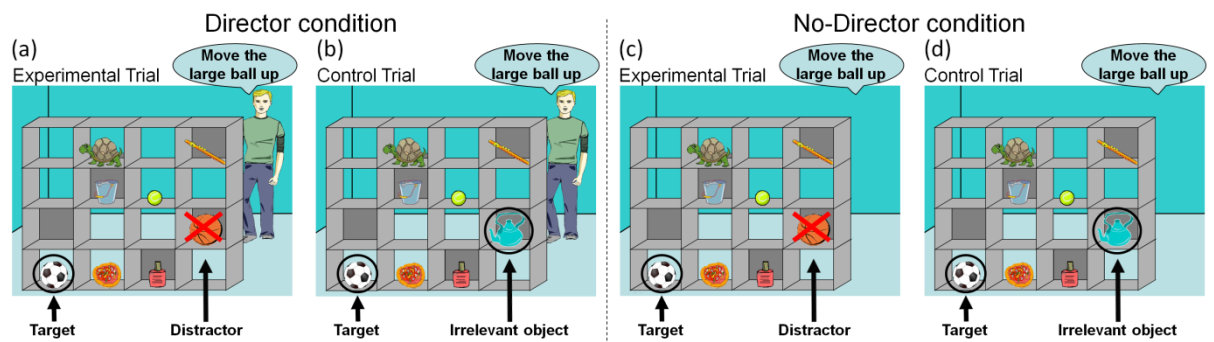
4 set of shelves containing eight different objects and were asked to move one of the objects in each trial.



**Figure 1** Instruction screens used to explain the task to participants. The set of shelves on the left is shown from the participant's point of view and the set of shelves on the right is shown from the director's point of view.

In the Director condition a 'director' was shown standing behind the shelves, viewing the shelves from behind (**Figure 1**). Participants were asked to listen to the director's instructions to move one of the objects (e.g. "Move the large ball up") and respond. Participants were told they should take the director's viewpoint into account when following his instructions. They were told that objects in slots with a grey background were only visible to them, while the other objects could be seen from either side of the shelves. In Experimental trials, the instruction referred to one object ('target') given the director's point of view, but would refer to another object ('distractor') if one assumes the participant's perspective (**Figure 2a**). As such, participants must take the director's perspective into account in order to respond correctly in an Experimental trial. In Control trials the distractor object was replaced by an irrelevant object and the instruction referred to an object that was visible to both the participant and the director (**Figure 2b**). In Filler trials instructions referred to single objects that were visible to both the participant and the director (e.g. the turtle in **Figure 2**).





**Figure 2** Stimuli of the perspective taking task. (a-b) Director condition. The participant heard the instruction “Move the large ball up” from the director. In an Experimental trial (a), if the participant did not take the director’s perspective into account he/she would move the basketball instead of the football, which is the second largest and can be seen by both the participant and the director. In a Control trial (b), the distractor is replaced by an irrelevant object. (c-d) No-Director condition. The participant is told that instructions do not refer to items in slots with grey backgrounds, the correct responses are therefore the same as in the Director condition.

In the No-Director condition, the only difference was that the director was removed from the stimuli (**Figure 2c-d**). Participants were told that the auditory instructions would only refer to items in clear slots and not items that were on slots with a grey background.

Forty-eight pairs of shelf configurations were created as Experimental and Control trials. All shelf configurations depicted eight objects, and included either three (Experimental trials) or two (Control trials) exemplars of the same object which differed in position (top/bottom) or size (large/small) (see **Appendix I**). In Experimental trials the exemplars were distributed such that the distractor object (the topmost, bottommost, smallest or largest object) was in a grey background slot, while the target object (the second topmost, bottommost, smallest or largest object) and the third object (“object 2”) were in a clear slot (**Figure 2a,c**). In the Control trials the distractor was replaced by an irrelevant object (**Figure 2b, d**). The rest of the objects were unique objects distributed among three grey background slots and two clear slots. Another 48 pairs of shelf configurations were created for the Filler trials, in which the instructions referred to single objects that were on a clear slot. Taken together, the stimuli were divided such that half were presented in the Director condition and the other half in the No-Director condition (counterbalanced across participants), and each participant saw 12

Experimental trials, 12 Control trials, and 24 Filler trials in the Director and No-Director conditions respectively. The order of stimulus presentation was counterbalanced between participants.

The materials and design of the present study differed from Dumontheil et al.'s (2010) study in three ways. First, we included a greater number of Experimental and Control trials to increase statistical power and allow eye-tracking data analysis. Second, while each shelf configuration was used with three successive instructions in Dumontheil et al.'s (2010) study, a different shelf configuration was used in each trial in the present study to minimise participants' learning strategies. Third, whereas participants had to pretend to drag the object in the previous study, they were able to click and drag any object on the grid in the present study.

#### *2.2.2.2 Inhibitory control tasks*

We used two different Go-NoGo tasks to measure participants' inhibitory control, both of which had one within-participant factor Trial Type (Go, NoGo). The Simple Go-NoGo task was based on the standard Go-NoGo paradigm (Simmonds, Pekar, & Mostofsky, 2008). A coloured square was presented on the left or right side of the screen in each trial. Participants had to indicate which side of the screen it appeared on if the square was green (a Go trial) and inhibit their response if the square was red (a NoGo trial). The Complex Go-NoGo task was identical except that it used yellow and blue squares and included a 1-back working memory (WM) requirement (see Simmonds et al., 2008), such that participants had to indicate on which side of the screen the square was shown (Go trials) except when a blue square was preceded by a yellow square (NoGo trials).

#### *2.2.3 Procedure*

Participants were tested individually in one session lasting approximately 45 minutes. They completed the various tasks in the following order: (1) the Director task (Director condition, then No-Director condition), (2) the inhibitory control tasks (Simple Go-NoGo task, then Complex Go-NoGo task) and (3) the vocabulary subtest of the WASI. Eye-movements during the Director task were recorded using a Tobii TX300 eye-tracker at a sampling rate of 60Hz. Stimuli for the Director task were presented using E-Prime 2

and the Go-NoGo tasks were programmed in Cogent ([www.vislab.ucl.ac.uk/Cogent/index.html](http://www.vislab.ucl.ac.uk/Cogent/index.html)) running in Matlab 7.0 (MathWorks).

Standardised instructions were read to participants prior to the Director task. For the Director condition, participants were told that the director had a different view of the shelves (**Figure 1**) and that his point of view must be considered when following his instructions to move objects. We asked each participant to give an example of an object that both he/she and the director could see, as well as an object that the director could not see to ensure that they understood the task. For the No-Director condition, participants were told that the director was no longer present and that instructions would only refer to items on clear slots, such that they should ignore items in slots with grey backgrounds when performing the task. The Director condition was always presented prior to the No-Director condition in order to prevent participants from applying the rule provided in the No-Director condition (Dumontheil et al., 2010).<sup>1</sup> Participants were presented with two filler practice trials before each condition.

In each trial the visual stimulus and auditory instructions were presented over a period of 2.2 s. The visual stimulus remained on the screen for another 3.8 s, during which participants responded by clicking an object and dragging it to a different position. A response was considered correct if the target object had been selected. Response times (RT) were measured from the onset of the display.

In the Go-NoGo tasks, each square was presented for 400 ms, with a 600-800 ms jittered fixation cross intertrial interval (ITI). Participants responded by pressing the left or right key using their right index or middle finger respectively (adapted from Watanabe et al., 2002). Participants performed two practice blocks prior to each task: the first practice block presented ten Go trials to establish a habitual response; the second one

---

<sup>1</sup> Note that the Director condition was always presented prior to the No-Director condition for the reasons stated and it is how it was done in the original study (Dumontheil et al., 2010). I acknowledge that this means that the conditions were not counterbalanced and that while the argument is that the rule provided in the No-Director condition can alter participants' performance in the Director condition, one cannot be sure that the opposite is not the case, i.e. participants perform better in the No-Director task because of exposure to the Director condition. However, in a previous study (Apperly et al., 2010) with adults participants where these two conditions were performed as between-subject design, the results found were similar to the current ones in terms of error, in that participants performed significantly worse in the Director condition than in the No-Director condition.

included six Go and four NoGo trials. Practice was repeated if participants made two or more errors. Each participant performed 80 test trials on each task (25% NoGo).

## 2.2.4 Data Analysis

### 2.2.4.1 Behavioural data analyses

All analyses were processed in SPSS and R. Statistical significance was set at  $p < 0.05$ . Bonferroni-corrected post-hoc t-tests were performed to explore significant main effects and interactions further.

#### 2.2.4.1.1 Director task

Mixed repeated measures ANOVA with two within-participant factors (Condition and Trial type) and one between-participant factor (Age group: Young Adolescent, Older Adolescent, Adult) were performed on participants' mean accuracy (percent errors) and reaction times (RT). Data for Filler trials were not analysed; participants made fewer than 2% errors on these trials on average. As verbal IQ differed between groups, we conducted an additional mixed repeated measures ANOVA which included verbal IQ as a covariate.

#### 2.2.4.1.2 Inhibitory control tasks

For both the Simple and Complex Go-NoGo tasks, mean accuracy (percent errors) was calculated for each participant in each Trial type (Go, NoGo) and median RT was calculated for correct Go trials. A 2 x 3 mixed ANOVA with a within-participant factor Trial type (Go, NoGo) and a between-participant factor Age group (Young Adolescent, Older Adolescent, Adult) as was performed for each task. One-way ANOVAs were performed to examine the effect of Age group on participants' RTs in Go trials in each task.

#### 2.2.4.1.3 Association between the Director and Inhibitory control tasks

Regression analyses were performed to investigate whether age-related changes in performance during the DT were associated with performance on the inhibitory control tasks. The difference in percentage errors between Director and No-Director Experimental trials, which is the critical measure of interest, was entered as the

dependent variable. In a first step, age was entered as continuous variable. In a second step, the four percentage-error measures of the Simple and Complex Go and NoGo trials were entered in a stepwise regression. Finally, in a third step, IQ was entered to assess whether it accounted for variance associated with age in this model.

#### *2.2.4.2 Eye-tracking data analyses*

Eye movements were analysed by computing a target advantage score which is the average probability of looks at the target object minus the average probability of looks at the distractor (Kronmüller & Barr, 2015). A look to the target was defined as a look to the object or the slot the object was on. Likewise, a look to the distractor (or irrelevant object in Control trials) was defined as a look to the object or the slot the object was on. Note that since the “distractor” was an irrelevant object in the Control trials, target advantage scores in the Control trials represented the difference between looks to the target and looks to the irrelevant object.

Target advantage scores were calculated over 50ms time-bins for the figures where 0 ms is the noun onset. For statistical analyses the target advantage scores were calculated for five different time windows (regions) during a trial. The five regions were time-locked to the onset of words in the auditory stimuli. Region 1 was the verb (“move”), Region 2 was the article (“the”), Region 3 was the modifier (“large”), Region 4 was the critical noun region (“ball”) and Region 5 was the directional preposition (“up”). Analyses of data in these regions were offset by 200ms to account for the time required for planning and launching an eye movement (Hallett, 1978). Data in each region were analysed for the Director and No-Director tasks separately. Since we did not have predictions regarding participants’ eye-gaze prior to the modifier (e.g., “large”), we focused our analyses on data in Regions 3 to 5 only.

First, 2 (Condition) x 2 (Trial type) x 3 (Age group) mixed ANOVAs were performed on correct trials for each region. Data from two participants from the Young Adolescent group were excluded from this analysis as they did not have enough correct trials. Second, 2 (Accuracy) x 3 (Age group) mixed ANOVAs were performed separately for correct and incorrect Director Experimental trials, the key trials of interest, for each region. As both adolescent groups did not have many correct trials (in some cases only

one or two), and some adults did not have incorrect trials, some participants (five young adolescents, five older adolescents and nine adults) had to be omitted from this analysis due to the lack of eye gaze data. Given the clear differences in accuracy between groups, we believe that it was not suitable to combine correct and incorrect trials in the analysis. However, the results of such analyses are available in **Appendix I**.

## 2.3 RESULTS

### 2.3.1 Behavioural results

#### 2.3.1.1 Director task

##### 2.3.1.1.1 Accuracy

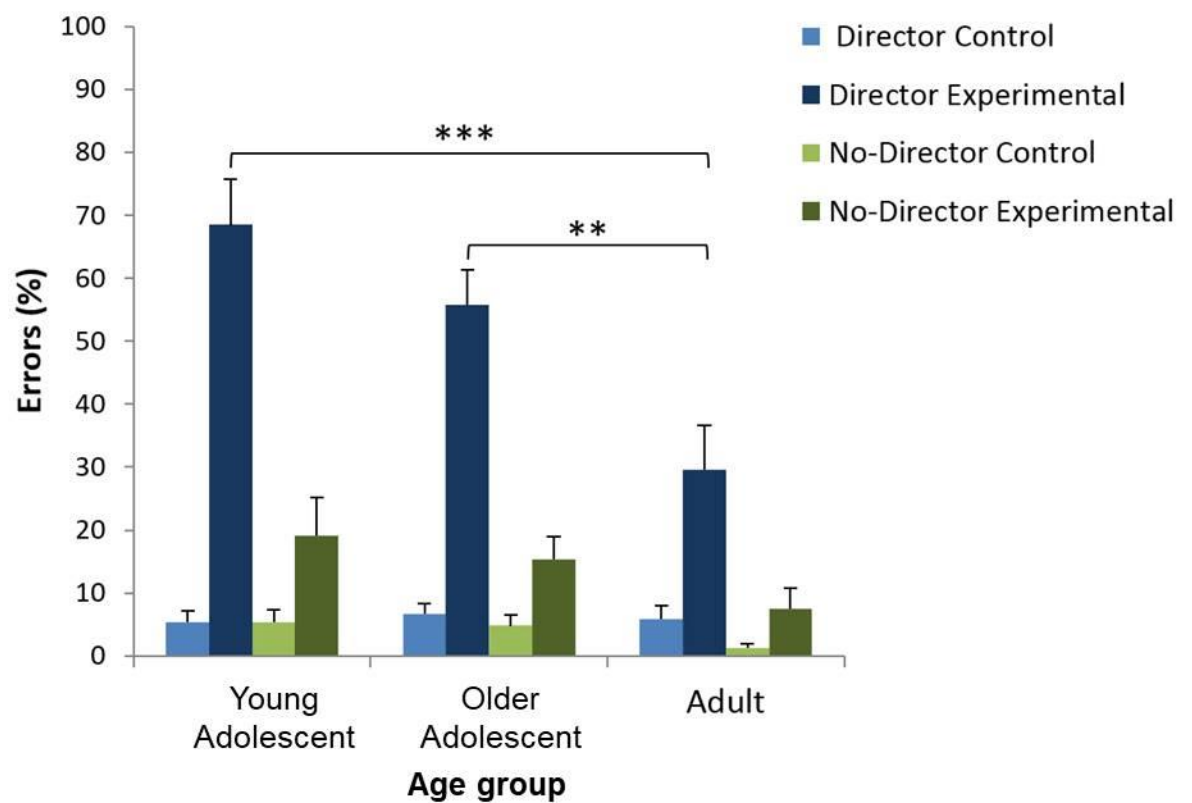
**Figure 3** shows participants' percentage error as a function of Age group and **Table 1** shows the results of the statistical analysis. There were significant main effects of Condition ( $\eta_p^2 = .591$ ), Trial type ( $\eta_p^2 = .675$ ), and Age group ( $\eta_p^2 = .220$ ), which were qualified by significant Age group x Trial type and Age group x Condition two-way interactions, and a significant three-way interaction ( $p = .013$ ,  $\eta_p^2 = .142$ ). We explored this interaction further by analysing the Experimental and Control trials separately. No significant effects were found in the Control trials (all  $ps > .08$ ). In contrast, in Experimental trials, there was a main effect of Condition ( $F(1,57) = 83.57$ ,  $p < .001$ ,  $\eta_p^2 = .595$ ), a main effect of Age group ( $F(2,57) = 8.10$ ,  $p = .001$ ,  $\eta_p^2 = .221$ ), and a significant interaction between Condition and Age group ( $F(2,57) = 3.67$ ,  $p = .032$ ,  $\eta_p^2 = .114$ ). Follow-up analysis examining Experimental trials in the Director and No-Director tasks separately revealed a significant effect of Age group in Director Experimental trials ( $F(2,57) = 8.10$ ,  $p = .001$ ,  $\eta_p^2 = .221$ ) but not in No-Director Experimental trials ( $F(2,57) = 1.85$ ,  $p = .167$ ). Follow up t-tests for the Director Experimental trials showed that both the younger and older adolescent groups made significantly more errors than adults (young adolescent:  $t(32) = 3.73$ ,  $p = .001$ ,  $d = 1.321$ ; older adolescent:  $t(44) = 2.94$ ,  $p = .005$ ,  $d = .887$ ), but there was no difference between the adolescent groups ( $t(38) = 1.36$ ,  $p = .183$ ).

**Table 1** Omnibus repeated measures ANOVA of percentage error in the Director task.

	df	F
Omnibus ANOVA		
Age group	2, 57	8.06**
Condition	1, 57	82.27**
Trial type	1, 57	118.49**
Age group x Condition	2, 57	2.27
Age group x Trial type	2, 57	6.65**
Condition x Trial type	1, 57	70.73**
Age group x Condition x Trial	2, 57	4.73*

\*\* $p < .01$

\* $p < .05$



**Figure 3** Percentage error (mean + SE) in the Director task for each age group.

\*\* $p < .01$ ; \*\*\* $p \leq .001$ .

As the age groups significantly differed in terms of their verbal IQ, we repeated the 2 x 2 x 3 mixed repeated measures ANOVA with standardised IQ (z-score) as a covariate. The same main effects and interactions were found.

To summarise, younger and older adolescents made more errors than adults only in the Director Experimental trials, which required taking the perspective of the Director into account (Director Experimental trials), in contrast to the rule-based control condition (No-Director Experimental trials). This replicated the behavioural results reported in Dumontheil et al. (2010).

#### *2.3.1.1.2 Response time*

Analysis of RT showed that the younger adolescents responded more slowly than the older adolescents and adults in all trial types except for the Director Experimental trials, which showed no difference between Age groups (see **Appendix I** for analyses). These results are also in line with those reported in Dumontheil and colleagues (2010).

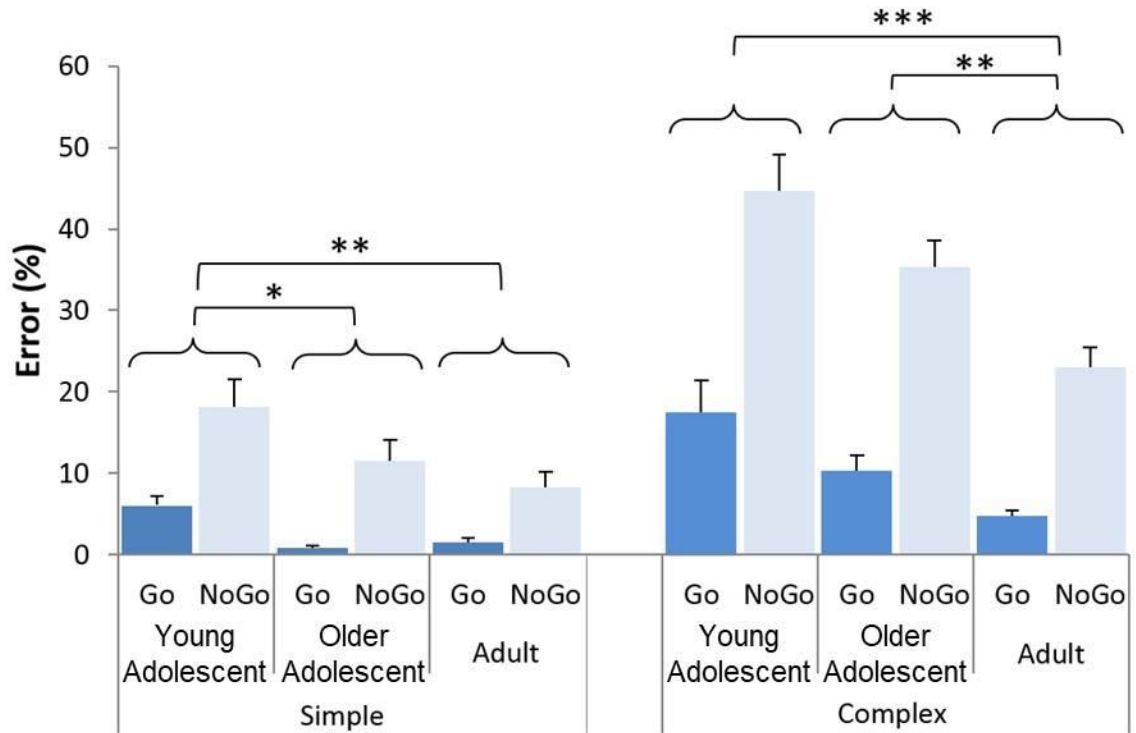
#### *2.3.1.2 Inhibitory control tasks*

**Figure 4** shows participants' percentage error in the Simple and Complex Go-NoGo tasks.

##### *2.3.1.2.1 Simple Go-NoGo*

The repeated measures ANOVA revealed significant main effects of Trial type ( $F(1,57) = 50.34, p < .001, \eta_p^2 = .469$ ) and Age group ( $F(2,57) = 5.29, p = .008, \eta_p^2 = .156$ ) showing that more errors were made in NoGo trials than Go trials, and that the younger adolescent group made more errors than the older adolescent ( $p = .027$ ) and adult groups ( $ps < .03$ ), who did not differ ( $p = 1$ ). The interaction between Age group and Trial type was not significant ( $p > .2$ ).





**Figure 4** Percentage error (mean + SE) in Go and NoGo trials in the Simple and Complex inhibitory control task for each age group. \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p \leq .001$ .

#### 2.3.1.2.2 Complex Go-NoGo

We found significant main effects of Trial type ( $F(1,57) = 108.87$ ,  $p < .001$ ,  $\eta_p^2 = .656$ ) and Age group ( $F(2,57) = 11.44$ ,  $p < .001$ ,  $\eta_p^2 = .286$ ) in the Complex Go-NoGo task, and no interaction between these factors ( $p > .10$ ). Here, the younger and older adolescent groups did not differ ( $p > .10$ ), but they both made more errors than the adult group ( $ps < .01$ ).

#### 2.3.1.2.3 Simple and Complex Go RT

Analyses of median RT in correct Go trials revealed that, in both the Simple and Complex Go-NoGo tasks, the young adolescent group was significantly slower than the older adolescent and adult groups, who did not differ from each other (See **Appendix I** for RT analyses).

#### 2.3.1.2.4 Association between Inhibitory control tasks and the Director task

Regression analyses showed that age significantly accounted for 14.4% of variance in the difference in percentage error between Director and No-Director Experimental trials (**Table 2**). Inhibitory control, as measured by the Simple NoGo percentage error, accounted for an additional 9.2% of variance, with more errors on Simple NoGo trials predicting more errors in Director vs. No-Director Experimental trials. The three other measures of Go-NoGo percentage error were not significant predictors (all  $\beta$ s < .124 and  $p$ s > .370). The effect of age was still significant in the second model, suggesting partially independent effects of age and inhibitory control. Finally, entering verbal IQ as an additional regressor did not improve the model and both the independent effects of age and Simple NoGo percentage error remained significant (**Table 2**).

**Table 2** Multiple regression analyses of Director task performance. The difference in percentage error between Director and No-Director Experimental trials was entered as the dependent variable.

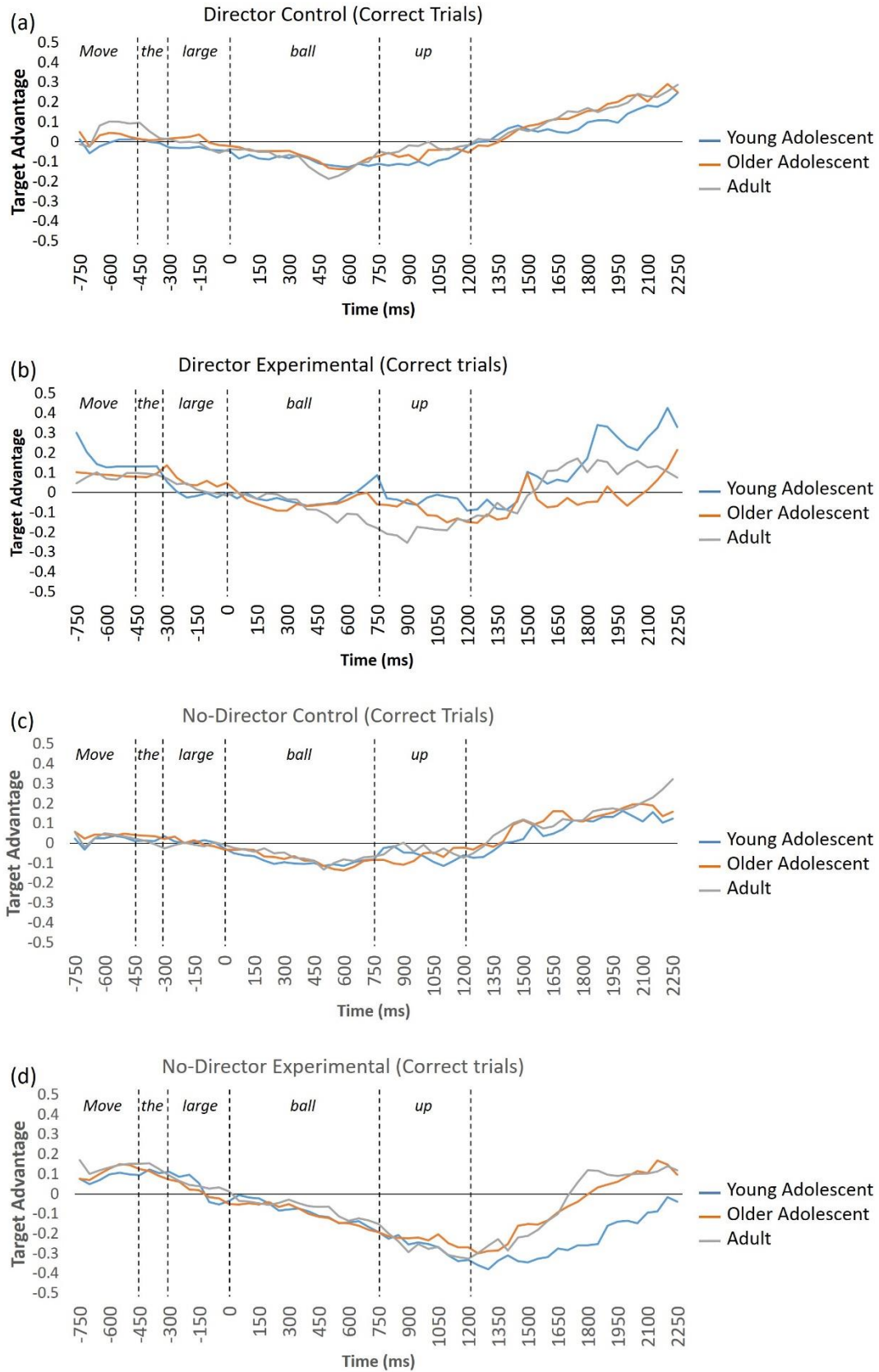
		$\beta$	t	p
<b>Model 1</b> $R^2 = 14.4\%$ ; $F(1,57) = 9.60$ , $p = .003$	(Constant)		5.421	.000
	Age	-.380	-3.099	.003
<b>Model 2</b> $R^2 = 23.6\%$ ; $F(2,56) = 8.64$ , $p = .001$ $\Delta R^2 = 9.2\%$ ; $F(1,56) = 6.72$ , $p = .012$	(Constant)		5.071	.000
	Age	-.298	-2.464	.017
	Simple NoGo	.314	2.591	.012
<b>Model 3</b> $R^2 = 25\%$ ; $F(3,55) = 6.11$ , $p = .001$ $\Delta R^2 = 1.5\%$ ; $F(1,55) = 1.09$ , $p = .302$	(Constant)		3.358	.001
	Age	-.287	-2.360	.022
	Simple NoGo	.311	2.575	.013
	Verbal IQ	-.122	-1.043	.302

#### 2.3.2 Eye-tracking results

**Table 3** shows mean target advantage scores across conditions and **Table 4** shows the results of the statistical analyses. **Figure 5a-d** show plots of average target

advantage over time. No significant effects were found in Region 3 (“large”; all  $ps > .2$ ). Main effects of Condition and Trial type were marginal in Region 4 (“ball”) ( $ps < .10$ ) and significant in Region 5 (“up”) (Condition:  $\eta_p^2 = .042$ ; Trial type:  $\eta_p^2 = .234$ ). The Condition  $\times$  Trial type interaction was significant in both Regions 4 ( $\eta_p^2 = .031$ ) and 5 ( $\eta_p^2 = .035$ ). Follow-up analyses revealed no significant effects in Control trials in either region ( $ps > .3$ ), while the target advantage in Experimental trials was significantly smaller in the No-Director task than in the Director task in both Regions 4 ( $F(1, 57) = 8.22, p < .01, \eta_p^2 = .055$ ) and 5 ( $F(1, 57) = 10.84, p = .002, \eta_p^2 = .090$ ). No significant effects involving Age group was observed in any of the regions (all  $ps > .15$ ).

To summarise, in correct trials, participants showed a smaller target advantage in No-Director Experimental trials than in Director Experimental trials upon hearing the noun. Crucially, their eye-movement patterns did not vary across age groups in any of the regions.



**Figure 5** Target advantage scores (looks to the target minus looks to the distractor (or irrelevant object for control trials)) per condition over 50ms time bins.

**Table 3** Mean target advantage scores (SE) for correct trials in each condition for each region. Standard errors are shown in parenthesis

Condition & Trial type	Region 3 "large"	Region 4 "ball"	Region 5 "up"
Young Adolescent			
Director Control	-0.08 (.02)	-0.12 (.02)	-0.02 (.04)
Director Experimental	0.00 (.03)	-0.04 (.03)	-0.07 (.07)
No-Director Control	-0.04 (.02)	-0.08 (.02)	-0.05 (.03)
No-Director Experimental	-0.03 (.04)	-0.15 (.04)	-0.28 (.06)
Older Adolescent			
Director Control	-0.03 (.02)	-0.09 (.02)	0.00 (.02)
Director Experimental	-0.03 (.02)	-0.07 (.04)	-0.11 (.05)
No-Director Control	-0.03 (.01)	-0.10 (.02)	0.01 (.02)
No-Director Experimental	-0.05 (.02)	-0.16 (.02)	-0.24 (.03)
Adult			
Director Control	-0.05 (.02)	-0.09 (.02)	0.03 (.02)
Director Experimental	-0.01 (.03)	-0.12 (.03)	-0.15 (.05)
No-Director Control	-0.02 (.02)	-0.07 (.02)	0.01 (.03)
No-Director Experimental	-0.01 (.03)	-0.13 (.04)	-0.24 (.05)

**Table 4** Omnibus repeated measures ANOVA of target advantage for correct trials in each region.

	df	Region 3 "large"	Region 4 "ball"	Region 5 "up"
Omnibus ANOVA				
Age group	2, 55	0.44	0.13	0.20
Condition	1, 55	0.00	3.73^	8.52**
Trial type	1, 55	0.77	3.29^	81.58**
Age group x Condition	2, 55	0.74	1.51	0.64
Age group x Trial type	2, 55	1.14	1.07	1.05
Condition x Trial type	1, 55	1.43	11.45**	8.12**
Age group x Condition x Trial	2, 55	0.24	1.96	0.42

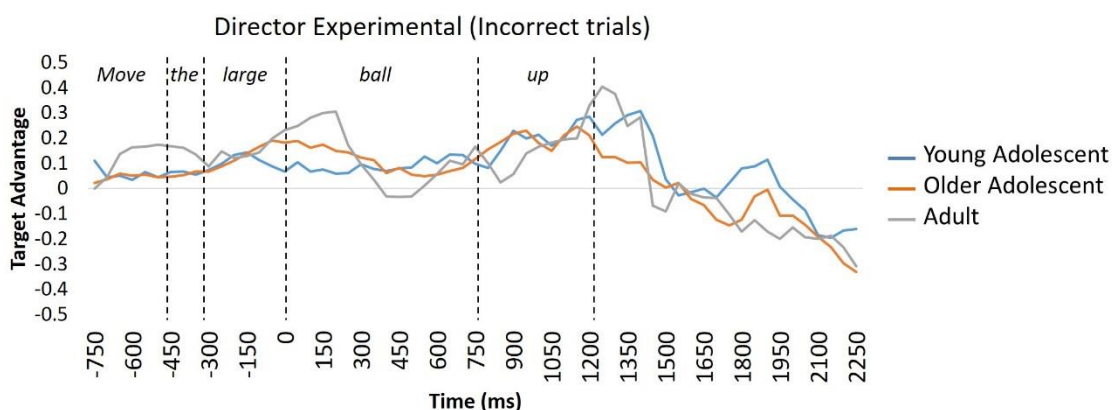
\*\* $p < .01$

\* $p < .05$

^ $p < .1$

We conducted additional analyses to examine the differences in eye-gaze between correct and incorrect Director Experimental trials, the key trials of interest in the Director paradigm (**Table 5**). We included an additional region (pre-response region) to check if participants looked at the object they chose. This region was defined as the 200ms before response time in a given trial. Graphs of target advantage indicate again a similar pattern across age groups (**Figure 5b and 6**). Mixed 2 (Accuracy) x 3 (Age group) ANOVAs revealed a main effect of Accuracy across Regions 3 to 5 and the Pre-response region (Region 3:  $\eta_p^2 = .240$ ; Region 4:  $\eta_p^2 = .271$ ; Region 5:  $\eta_p^2 = .224$ ; Pre-response:  $\eta_p^2 = .118$ ). Participants had a bigger target advantage in incorrect trials than correct trials in Regions 3 to 5, and, conversely, a bigger target advantage in correct than incorrect trials in the pre-response region. No significant effects involving Age group were observed (all  $ps > .3$ ).

To summarise, analyses of eye-movement data in the Director Experimental trials revealed clear differences between correct and incorrect trials, but there was no significant effects involving Age group in any of the regions. In other words, there were no significant age-related differences in participants' eye-gaze patterns in correct as well as incorrect Director Experimental trials. Further, the effect of accuracy observed in Regions 3, 4, and 5 was reversed in the pre-response region, such that participants initially showed a greater target advantage in the incorrect trials than in the correct trials, and it was only right before they responded that they showed a greater target advantage in correct trials than in incorrect trials.



**Figure 6** Target advantage scores in each age group over 50ms time bins for incorrect Director Experimental trials.

**Table 5** Mean target advantage scores (SE) for correct and incorrect Director Experimental trials for each region.

Accuracy	Region 3 "large"	Region 4 "ball"	Region 5 "up"	Pre-response Region
Young Adolescent				
Correct	0.00 (.04)	-0.03 (.03)	-0.01 (.06)	0.32 (.15)
Incorrect	0.10 (.04)	0.11 (.03)	0.13 (.09)	-0.16 (.12)
Older Adolescent				
Correct	-0.03 (.02)	-0.09 (.04)	-0.14 (.05)	0.11 (.09)
Incorrect	0.17 (.05)	0.17 (.05)	0.19 (.06)	-0.12 (.11)
Adult				
Correct	0.00 (.04)	-0.09 (.03)	-0.07 (.08)	0.29 (.11)
Incorrect	0.27 (.06)	0.08 (.05)	0.18 (.11)	-0.01 (.12)

**Table 6** Omnibus repeated measures ANOVA F-values of target advantage for correct and incorrect Director Experimental trials in each region.

	df	Region 3	Region 4	Region 5	Pre- response
Omnibus ANOVA					
Age group	2, 38	1.74	0.54	0.12	0.80
Accuracy	1, 38	23.68**	30.23**	22.06**	12.53**
Age group x Accuracy	2, 38	1.03	1.02	0.87	0.70

\*\* $p < .01$

\* $p < .05$

## 2.4 DISCUSSION

In this study, we collected behavioural and eye-tracking data to investigate online use of ToM during perspective taking in young adolescents, older adolescents and adults. Experimental trials of the Director task required participants to take into account the director's perspective to determine the intended referent in the instructions (e.g., 'the large ball'). The No-Director condition required participants to follow an explicit avoidance rule to determine the correct referent. While both Director and No-Director conditions involved executive function, in particular inhibition, only the Director condition required an inference about the speaker's intentions given their ignorance of certain objects (Dumontheil et al., 2010).

The present study had three main objectives: (i) to replicate the behavioural findings of Dumontheil and colleagues (2010), (ii) to assess the role of inhibitory control in this task across age groups, and (iii) to compare the time course of information processing between young adolescents, older adolescents and adults through eye-tracking in order to determine in what ways the deployment of theory of mind differs across age groups.

#### *2.4.1.1 Behavioural data*

In the Director Experimental condition, the young and older adolescent groups performed worse than the adult group, but participants' accuracy did not differ across groups in the No-Director Experimental condition. These results are in line with Dumontheil et al.'s (2010) findings and provide further evidence for age-related differences between adolescents' and adults' tendency to take someone else's perspective into account. Conversely, response times decreased with age in all conditions except in Director Experimental trials, which did not vary with age, and is again similar to that observed by Dumontheil et al. (2010).

##### *2.4.1.1.1 Inhibitory control*

Results revealed age-related differences in participants' performance in both inhibitory control tasks. In the Simple Go-NoGo task the young adolescent group made significantly more errors and was significantly slower than the older adolescent and adult groups. These results are in line with previous studies that suggest that with a low-level cognitive load inhibitory control performance reaches adult level performance by approximately age 14 (Leon-Carrion, Garcia-Orza, & Perez-Santamaria, 2004; Lamm, Zelazo & Lewis, 2006; Luna et al., 2004). In the complex task, the young and older adolescent group performed worse than the adult group. Since the Complex task included a 1-back Working Memory (WM) requirement, these results are also in line with studies investigating working memory and have found age-related differences between adolescents and young adults (Conklin et al., 2007; Luna, Garver, Urban, Lazar, & Sweeney, 2004a).



#### *2.4.1.1.2 Association between Inhibitory control tasks and the Director task*

What was novel in this study was the investigation of the relationship between the inhibitory control tasks and the Director task. The results show that inhibitory control as measured by the Simple NoGo accuracy accounted for some of the variance in accuracy difference between Director and No-Director Experimental trials. Critically, inhibitory control accounted only for some of the variance due to age, as age remained a significant predictor, suggesting that some additional factors are behind age-related changes. These results are in line with a previous study by Vetter and colleagues (2013), who found that 15% of the variance of affective ToM performance was uniquely explained by age, indicating independent effects of age and inhibition.

Interestingly, no relationship was found between participants' performance in the Complex Go-NoGo task and the Director task. This is surprising as performance in the Complex Go-NoGo task showed age-related differences between the adolescent and adult group. This could be because the Complex task placed greater working memory demands than the Director task. A study that explored the relationship between working memory and perspective taking (Lin et al., 2010), found that participants with lower WM capacity performed more poorly on the Director task than participants with greater WM capacity, and that participants' performance in the Director task was worse during high WM load than low WM load trials. Therefore, WM can also be a factor. However, the Lin et al. (2010) study did not include a No-Director condition with matched general executive function demands, while the present study used the difference in accuracy between Director Experimental and No-Director Experimental trials as the measure of interest. In addition, there is not association with performance in the DT and the IC task that involves WM (complex Go-NoGo). Nevertheless, although the No Director condition is matched for EF and participants perform well in that task, as with inhibitory control however, I cannot eliminate the possibility that WM might be a factor that accounts for some of the variance in performance. Future research is required to shed light on this by using separate measures of WM and inhibition (Vetter, Altgassen, et al., 2013).

A possible limitation for the interpretation of the behavioural results is the difference in verbal IQ between age groups. However, the significant interactions with age observed in the Director task were still present when verbal IQ was added as a

covariate. Moreover, adding verbal IQ as a predictor in the multiple regression did not affect the results, as both the independent effects of age and Simple NoGo accuracy remained significant.

#### *2.4.1.2 Eye-tracking data*

Through eye-tracking we were able to investigate additional underlying factors in Director task performance. Examining correct and incorrect trials separately indicated that adults, and both adolescent groups do not differ in their online processing of the task (**Figure 5**). Analyses of Director Experimental trials data showed opposite effects of accuracy in the earlier regions (Regions 3, 4, and 5) and the pre-response region, such that participants initially showed a greater target advantage in the incorrect trials than in the correct trials. They only showed a greater target advantage in correct trials than in incorrect trials right before they responded. These results seem inconsistent with the perspective-adjustment model (Keysar et al., 2000, 2003) according to which one would expect that on both correct and incorrect trials, participants first attend to the distractor, the best-fit referent from an egocentric perspective. On correct trials participants would adjust to considering the director's perspective and attend to the target, while on incorrect trials, the second adjustment process would fail, as it is costly, and participants would remain focused on the distractor. However, as evident from the incorrect trials for all age groups, participants appear to consider the target early on in the trial and then, before responding, their eye gaze shifts to the distractor.

Based on the eye tracking results it seems that at the point when the Director indicates which object should be moved (Region 4), participants on correct trials are already on path to correctly considering the director's perspective. However, participants' strategy seems to be to first look at the objects they are not going to choose (or the objects they should not pick) as a process of elimination before focusing on the object that they will ultimately choose. This pattern has not been reported before. Other studies, such as Hanna et al. (2002), Heller et al. (2008) and Barr (2008) showed that bias in participants' eye gaze builds steadily toward the target after initial interference from the distractor. What sets this study apart from these eye-tracking studies is the fact that the distractor is the best fit for the description, making the task of ignoring the privileged object particularly challenging. Additionally, the description itself contains a relational

modifier ('big', 'bottom') that implicitly refers to a contrast set. Given these two factors, it should not be surprising that participants adopt a strategy of checking all objects of the same type (e.g. all balls on the display) to ensure they choose the correct one. Our eye-gaze data of object 2 (e.g. the second commonly viewable ball) suggests this also. It shows that prior to focusing on the object they chose, participants pay similar attention to both of the other two objects they eliminate (see Appendix I; **Figures S3, S4, S5**). The only other studies that used a similar set-up to ours are those reported in Keysar et al. (2000) and Wu & Keysar (2007). Neither of these papers reported eye-gaze data in full but their results are consistent with ours in that they found overall more looks to the distractor and first looks to the distractor were earlier than first looks to the target.

Adults, older adolescents and younger adolescents in the present study seemed to follow a process of elimination strategy in control trials as well, where only the two commonly viewable objects denoted by the noun (e.g. 'ball') were present and only one fitted the full description ('the large ball'). I propose that participants focus first on the second object to exclude it before looking at the target (see Appendix I; **Figure S3, S4, S5, S6**). Therefore, there may be a consistent strategy across all conditions, and all age groups, of looking at the objects that are not going to be chosen prior to focusing on the object to be picked.

If we accept that participants adopt a process-of-elimination strategy, then our results suggest that eye-gaze pattern on correct trials are influenced by the actual beliefs of the director early on, as we see the pattern emerge in Region 4 when participants process the modifying adjective ('large'). The idea that eye-gaze data reflects an influence of the speaker's perspective at an early stage is in line with a number of previous perspective-taking studies mentioned above (Hanna et al., 2003; Brown-schmidt, Gunlogson, & Tanenhaus, 2008; Heller et al., 2008). The results are also consistent with Nadig and Sedivy's (2002) observation that 5 and 6 year-old children showed early sensitivity to speaker's perspective in a much simplified director task. These papers proposed an alternative to the Perspective-Adjustment model, claiming that mental state information, like any other relevant information, is potentially available to be integrated in referential decision processes from the outset. On this Constraint-Based view, the extent to which mental perspective information is utilized

depends on the extent to which other constraints are conflicting and how salient or available the perspective information is (Brown-Schmidt & Hanna, 2011; Samson et al., 2010). Hanna and colleagues (2003) argued that the Director paradigm used in Keysar and colleagues (2000) and in this paper makes conflicting cues related to the linguistic form particularly strong, since the occluded object (the distractor) is in fact the best fit for the description. Other studies such as Brown-Schmidt (2009b) have suggested that varying the strength (or quality) of cues to the speaker's mental perspective can affect on-line referential processes. If varying cues to mental perspective can have an impact on ToM integration, it seems plausible from this Constraint-Based perspective that individuals may differ in the extent to which their referential processes exploit a given cue. Thus, an explanation of the age-related differences that are not accounted for by inhibitory control may lie in differences in participants' sensitivity of on-line processes to mental-perspective information. This is a hypothesis that requires further exploration.

If we assume that accuracy differences in our Director task are a product of varying abilities to integrate perspective information in incremental referential processes, rather than at a later, corrective stage, then we can make sense of the reaction time results reported here and in Dumontheil et al. (2010). These results showed shorter reaction times in the Director task compared to the No-Director task and also no RT difference in the Director condition between age groups. The observation made about these results is that the Director task engages a more efficient or rapid process than simple explicit rule following and does so to the same extent across age groups.

Taken together, the present results provide evidence for age-related differences between adolescents and adults in their online use of ToM. Contrary to perspective-adjustment accounts of the Director task, the current results suggest that all age groups appear to engage in the same kind of online processes during perspective taking but differ in how often mental state information informs incremental decision processes. Taking our results in the wider context of research into on-line use of Theory of Mind, I propose that one possible source of the difference between age-groups is their sensitivity to available cues to mental state information.

## 3 CHAPTER 3: FALSE-BELIEFS AND ADOLESCENTS

---

### 3.1 INTRODUCTION

In Chapter 2 I investigated the development of ToM use through eye-tracking in young adolescents (9-13 years old), as well as older adolescents (14-17 years old), and adults (19-29 years old), using a computerised version of the Director task (Keysar et al., 2000). The results provided evidence for prolonged development of online use of ToM in adolescents, and while we found that inhibitory control might be factor to the poorer performance, it only accounted for some of the variance. This result calls for further research into online ToM processes during this period of life. The current chapter uses a false belief task to address why adolescents fail more often in the specific perspective-taking task than adults.

The false-belief (FB) task is one of the main paradigms to assess ToM, the most prominent being the Sally-Anne task (Baron-Cohen et al., 1985). In such a task, participants are introduced to two characters: Sally, who is playing with a marble, and Anne, who is watching. Sally places the marble into her basket and leaves. While she is absent Anne takes the marble out of that basket and puts it into her box. Participants are then asked where Sally will look for the marble when she returns, requiring them to infer Sally's FB in order to respond correctly. Conventional false-belief tasks have been fundamental in examining ToM in preschool children. However, these studies have shown that children's performance reaches ceiling at around the age of 5 (Surian et al., 2007; Wellman et al., 2001; Wimmer & Perner, 1983). As such, few studies have investigated ToM in middle childhood and adolescence using age appropriate tasks and these have focused more on emotion identification (Keulers, Evers, Stiers, & Jolles, 2010), affective ToM (Sebastian et al., 2012; Vetter, Altgassen, et al., 2013) and perspective-taking (Dumontheil, Apperly, et al., 2010; Symeonidou, Dumontheil, Chow, & Breheny, 2016).

Recently, however, false-belief paradigms have been developed that involve more sensitive implicit measures, such as reaction times (Apperly et al., 2006; Back & Apperly, 2010), brain responses (Ferguson, Cane, et al., 2015; Geangu et al., 2013; Meinhardt et al., 2012) and eye-tracking (Ferguson, Apperly, et al., 2015; Ferguson & Breheny, 2012; Rubio-Fernández & Glucksberg, 2012) where adults' performance is less-than-ceiling. Despite the creation of more advanced false-belief tasks only a few have investigated children beyond the middle childhood years into early adolescence and none that we know of have investigated older adolescents (14 and older).

Looking at studies that focused on younger adolescents, two studies have used tasks that involved participants' understanding of belief. The first is study by Devine & Hughes (2013) which, I reviewed in Chapter 1 but repeat here for the reader's convenience. The study involved a novel Theory of mind reasoning task - the Silent Film task - and tested children between 8-13 years of age. The task is a film-based analogue to the Strange stories task (Happé, 1994) that is well suited to exploring the ability to understand others' beliefs in children beyond the age of 7. The participants were presented with short film clips that depicted scenarios with deception, false belief, and misunderstanding and were tasked with explaining the behaviour of a character. Performance on the task improved with age, independently of individual differences in verbal ability and socioeconomic status, suggesting that ToM abilities do continue to develop through middle childhood.

More recently, Begeer and colleagues (2016) tested typically developing 7-13 year olds as well as children with autism spectrum disorder (ASD) using a modified false belief task- the Sandbox task- and a visual hindsight bias task. The modified Sandbox task is similar to a Sally-Anne task but instead of there being two possible locations, there is only one- a large sandbox- where the object is placed while the character is present and then reburied somewhere else in the sandbox while the character is absent (Begeer et al., 2016). Participants are then asked to point where in the box they think the returning character will look for the object. A memory control condition was also included where participants were asked to state where the character placed the object before he/she left. The authors argue that the Sandbox allows for a continuous response scale that measures egocentric bias as participants have to pick an area they think the

character will look for the object within the same location instead of picking one of two locations. Surprisingly, they found that both groups (with and without ASD) were more biased in the false-belief condition than in the memory control condition. According to the authors, this suggests that egocentric tendencies between these two groups are more similar than previously reported.

### 3.1.1 The current study

So far studies that investigated ToM abilities in middle childhood and adolescence point to a consensus that there are age related differences in ToM beyond childhood into adolescence. However, it is not clear what factors account for this age-related improvement. Results from explicit false belief tasks on young adolescents show prolonged ToM development (Begeer et al., 2016), and after controlling for verbal ability and family affluence (Devine & Hughes, 2013). The experiment reported in Chapter 2 (Symeonidou et al., 2016) showed differences in how often adolescents successfully take a speaker's perspective early on. If perspective-taking was successful, the time course of their attentional bias to the correct target did not differ from successful adults. Moreover, in unsuccessful trials the time course of gaze bias did not differ between adolescents and adults either. One possibility is that the poorer performance is due to high EF demands of the task rather than limited use of ToM in adolescence. Rubio-Fernandez (2016) argues that ToM is not necessary for successful performance on the DT task but rather selective attention alone would be sufficient (Rubio-Fernández, 2016). However, the results in Chapter 2 showed that while IC was a contributing factor, it only accounted for some of the variance in the difference in percentage error between Director and No-Director Experimental trials. Moreover, in the No Director condition, that matched the Director task in terms of EF including IC and WM, adolescents performed as well adults. Therefore, the results suggest that age-related differences on the DT are not simply due to high EF demands.

In Chapter 2 I suggested that, in line with theories of language processing, such as the Constraint-Based view, age-related differences might be due to how sensitive a participant is to available cues about mental states and how well they can integrate that information. For instance, Brown-Schmidt (2012) shows that performance on a DT is improved when the information in the experimental context better signals what is

common-ground information. According to the Constraint-Based view, mental state information is used depending on the extent to which other constraints are conflicting and how salient or available that information is in the context. The DT involves a linguistic conflict as the linguistic description always fits the competing object better than the target. In addition, the available cues (i.e. the presence of an avatar, the director, in every trial, and an image of the grid representing occlusion behind certain objects) do not provide very salient information about the distinct perspective of the director. The DT also involves complex integration processes. A participant has to use mental state information from available cues while at the same time inhibiting conflicting linguistic cues in order to ascertain someone else's belief.

An additional consideration may be that, in the DT, the participant then has to plan her own action based on someone else's incomplete belief very rapidly. Taken together, the task requires a rapid complex interplay of ToM and EF. The question then arises whether the age-related differences found in Chapter 2 are simply down to differences in ability to attend to relevant mental state information and integrate that information into the referential processes.

In order to test this possibility, we developed a task which we predict would be easier for adolescents to perform by making mental state information more salient, while keeping the level of ToM inferential complexity of the tasks at a comparable level. The current task examines inferences based on false beliefs, rather than ignorance, as in DT. Although somewhat different in nature, both tasks require participants to infer the speaker's intention based on information about the speaker's mental state, which is different from one's own. The aim is to determine whether adolescents demonstrate sensitivity to other peoples' beliefs as spontaneously and as early as adults in this case.

The task used in the current experiment is adapted from Ferguson and colleagues (2015) and involved videos portraying true and false-belief scenarios. Experimental videos presented two characters, Sarah and Jane. At the beginning of the video both Sarah and Jane are standing behind a table which has a target object (e.g. a pen) in the centre and three containers, one on the left, one in the middle and one on the right. While Jane is watching, Sarah moves the object into one of the three containers. Sarah then moves the object into one of the other containers, either while



Jane is still present and watching (true-belief for Jane) or after she has left (false-belief for Jane). Participants were then presented with a still image of the final scene of the video while they heard an auditory description such as, “Jane will look for the pen in the container on the [left/middle/right]”. In a TB condition the final container would be the object’s true location and in the FB condition it would be the initial container (thus, true according to Jane’s limited knowledge). During this auditory presentation participants’ eye-gaze around the visual scene (the three containers) was recorded and time-locked to the auditory stimulus. This showed which referent (i.e. which container) participants were considering/expecting over time. Ferguson and colleagues (2015) argued that listeners are given time to “set up expectations about forthcoming referents” before the onset of the disambiguating information by including the referent at the end of a longer linguistic stimulus (Ferguson, Apperly, et al., 2015, p. 52).

Results in Ferguson and colleagues (2015) revealed that adult participants had different visual biases for TB and FB conditions before they heard the disambiguating information, suggesting that they had inferred the character’s perspective. This effect was greater when an active task prompted early perspective-taking, but also when participants were simply passive observers. Ferguson and colleagues (2015) argue that this suggests adults are sensitive to other people’s perspective even when they do not have an explicit reason to attend to the other’s beliefs. This is in line with other studies that have found adults can track other people’s perspective even when they do not have a specific reason to do so (Cohen & German, 2009; Samson et al., 2010; Schneider, Bayliss, Becker, & Dux, 2012; Schneider & Low, 2016; Schneider, Slaughter, & Dux, 2015, 2017). However, as in Ferguson & Breheny (2012) results also showed that visual preference to the correct container in the FB condition was delayed in comparison to the TB conditions. In the FB condition they only showed a fixation bias for the initial container once auditory input confirmed which container it was. This suggests that it was harder for adults in the FB condition to use perspective information early on to generate expectations about the correct container.

In this chapter I want to investigate whether adolescents are just as sensitive as adults are to others’ beliefs even without being given an explicit instruction to track their mental states. I chose this particular FB task for the following reasons. Firstly, the visual

information in the FB task is less complex and more salient, as there are always only three possible consistent locations (left, middle, right) in every trial. Secondly, in comparison to the DT task, the extended linguistic stimulus prior to the critical word will allow us to track the online process of inferring other people's beliefs in adolescents over time. Thirdly, although the FB and the Director task both involve a reality location conflict that needs to be resolved, the integration complexity is less in the current task as one only has to predict someone else's action based on their false-belief rather than plan their own action. Moreover, the FB task does not involve a linguistic conflict as in the DT used in Chapter 2, (where the linguistic description always fits the competing object better than the target). Thus, the demands of integrating conflicting information online are lowered. Lastly, this task is appropriate for younger and older adolescents as adults themselves do not display a ceiling effect - finding it hard to integrate information early enough to reliably anticipate the correct container during a FB condition.

Overall, in comparison to the DT, the current task involves ToM inferences that are just as complex, but mental state information is more prominent and there is less conflicting information in the experimental context that increase the demands of integrating the relevant information. Thus, if the results found in Chapter 2 are due to rapid integration abilities of mental state information and EF then we would expect adolescents' performance in the FB task to be comparable with the adult group.

In addition to the main task we wanted to check if individual measures that have been shown to be associated with ToM differed between groups. Here we tested individual measures in executive function: inhibitory control (IC) and working memory (WM), as well as empathy. As presented in Chapter 1, previous studies have found an association between inhibition and the application of ToM inference in children (3-4 year olds in Carlson, Moses, & Claxton, 2004; 4-10 year olds in Hansen Lagattuta, Sayfan, & Harvey, 2014; 4-9 year olds in Lagattuta, Sayfan, & Blattman, 2010; 2-5 year olds in Nilsen & Graham, 2009; meta-analysis of 3-6 year olds in Devine and Hughes, 2014), in adolescents (Chapter 2; Symeonidou et al., 2016; Vetter, Altgassen, et al., 2013a) and in adults (Brown-Schmidt, 2009b). Previous research has also shown a relationship between working memory (WM) and perspective-taking (Bull et al., 2008; Cane et al., 2017; Lin et al., 2010; McKinnon & Moscovitch, 2007; Nilsen & Graham, 2009; Schneider,

Lam, et al., 2012), with a lower WM capacity or increased working memory load leading to poorer performance on ToM tasks.

Associations have also been found between empathy and ToM. As presented in Chapter 1, evidence from neuroimaging has suggested that ToM processes and empathy processes activate overlapping neuronal networks such as the medial prefrontal cortex, temporoparietal junction and temporal poles (Völlm et al., 2006). Moreover, in an ERP study Ferguson et al. (2015) found that participants with high levels of empathy are better at making inferences according to someone else's FB (which modulated the amplitude of the N400 effect), whereas people with lower levels of empathy were more likely to interpret events egocentrically. Behavioural data from Sebastian et al. (2012) showed that adolescents (11-16 years) had lower scores of affective empathy in comparison to adults (24-40) and importantly, adolescents also made significantly more errors on an affective ToM task where participants had to infer how a story character would react to another character's affective state, than adults.

In sum this chapter aims to i) replicate the adult findings from Ferguson et al. (2015), ii) examine whether adolescents can reliably infer others' (false) beliefs as spontaneously and early as adults, even without being given an explicit instruction to do so, iii) explore whether individual differences in inhibitory control, working memory or empathy are possible factors in constraining performance during online false-belief reasoning. For the adult data, we predict similar results as in Ferguson and colleagues (2015), namely that there should be a clear difference in looks to the containers between the TB and FB conditions early on - from when they hear whose perspective to take (i.e "Jane will look for the"). Moreover, there should be clear anticipatory looks to the final container in the TB condition, but weaker anticipation of the initial container in the FB condition. In contrast, according to previous data adolescents will be delayed at showing a differential pattern of looks to the containers between conditions, though the difference may still emerge before the disambiguating information. However, if the age-related differences found in Chapter 2 are due to limited abilities to rapidly integrate mental state information, then we could expect adolescents' performance in the FB task to be relatively better, possibly comparable with the adult group, as those factors are eliminated here. We also predict that at the point the location term (left, middle, right)

is presented auditorily both age groups should be more likely to look at the final container when Jane held a TB and more likely to look at the initial container when Jane held a FB about the object's location. In line with previous studies, we also expected to see age-related differences in IC, WM and empathy.

## **3.2 METHOD**

### **3.2.1 Participants**

Forty-one participants took part in this study, of which 17 were adults (24-36 years old,  $M = 26.9$ ,  $SD = 2.93$ ) and 24 were adolescents (11-18 years old,  $M = 15.53$ ,  $SD = 2.25$ ).<sup>2</sup> All participants were native English speakers. Adult participants were recruited from the UCL Psychology participant pool while adolescents were mostly recruited from London schools. Data from three adolescents were excluded from the analysis due to problems with the eye-tracking recording. Parents/guardians of all adolescent participants, as well as all adult participants were given information sheets prior to the study and informed consent was obtained. This study was approved by the UCL Research Ethics Committee.

### **3.2.2 Stimuli and design**

#### **3.2.2.1 False-belief task**

The task design and stimuli were based on a previous study by Ferguson, Apperly, Ahmad, Bindemann & Cane (2015). The experiment employed a 2 x 2 mixed design, with the Belief condition (true belief vs false belief) as the within-participant factor and Age group (Adult vs Adolescent) as the between-participants factor. There were sixteen sets of experimental videos and pictures with their respective auditory description in one of two conditions. The video clips were taken from the Ferguson and colleagues (2015) study which were recorded in a single session with two female actors. The auditory sentences were also from Ferguson and colleagues (2015) which were recorded from a female native British English speaker with a neutral intonation.

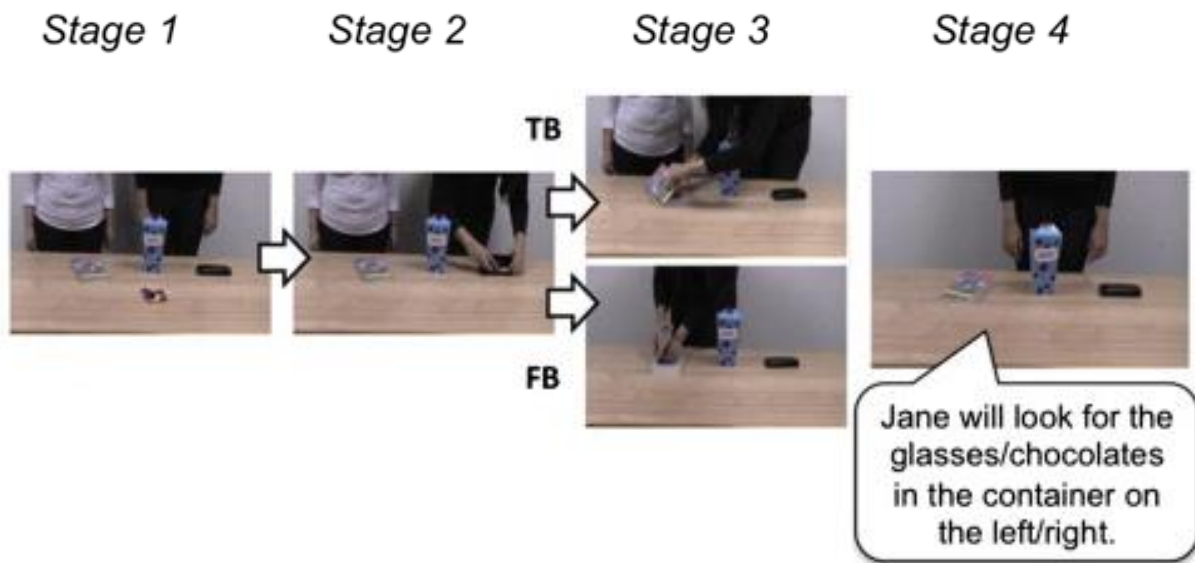
---

<sup>2</sup> Note that I piloted the task with 9 and 10 year olds in order to match age groups with Chapter 2. However, the task was too difficult for participants below the age of 11 as they were getting too tired and could not focus that long.

Two different video sequences set up the two Belief conditions (**Figure 7**). All videos involved a transfer event which began with the two actors (Jane and Sarah) standing behind a table. Videos began with three possible containers on the left, middle and right side of the table, and the target object in the centre of the table. In the first part of the video, Jane and Sarah were both present and Sarah moved the target object into one of the three containers while Jane watched. In the second part Sarah moved the target object into one of the other containers. Crucially, in one condition, Jane was present for this second transfer event which meant that Jane had a true belief about the object's location. However, in the other condition Jane left the scene after the first transfer events and was absent during the second transfer which meant that Jane has a false belief about the object's location. All videos concluded with Sarah standing alone behind the table and the three closed containers.<sup>3</sup> This final scene of the video was extracted and presented as a picture after the video ended. The spatial arrangements of the containers was counterbalanced to avoid viewing strategies by the participants.

---

<sup>3</sup> Note that the fact that only Sarah is present at the end of the videos and Jane is does not return, differs from the original Sally-Anne task, as Sally returns to the room before participants are asked where she will look for her marble. An argument therefore, can be made that there is a vagueness in time as it might not be as clear that one should consider what location she will search in on her return. However, the data from the original Ferguson et al., 2015 study show that adult participants in the active condition can anticipate the right location regardless suggesting that they interpret the sentence in the same way as if she were to have come back. In the passive condition, which is the same as the current study here, participants can anticipate the location in the TB condition.



**Figure 7** Visual displays presented to participants in the false-belief task. Stage one shows the 'start state' where both Jane and Sarah are standing behind the table. Stage two was the first part of the video where Sarah moves the target object in one of the three containers. In stage 3 the video showed Sarah moving the object into one of the other containers while Jane was present in true belief (TB) trials or after Jane had left in false belief (FB) trials. Lastly, in stage 4 a picture of the 'final state' from the video was presented to participants while they listened to the audio sentence.

While participants were presented with the picture that depicted that final state of the scenarios, they heard the following single-pre-recorded sentence, "Jane will look for the [object] in the container on the [left/middle/right]".<sup>4</sup> There were two different age groups (adult and adolescent) and both groups were told to look and listen to the videos.

One version of each item was assigned to one of two lists and participants were randomly assigned to one of these two lists. Each list contained sixteen unique experimental items, eight in each condition (true and false belief). Sixteen filler trials were also included in each list to prevent participants from guessing the aim of the study. The filler video scenarios depicted transfer events similar to those in the experimental trials. For half of the fillers, the second part of the video showed the target object being placed back into the same container that the object was transferred into during the first part of the video. This made it harder for participants to make predictions

<sup>4</sup> This description of the object's location prompts participants to assign left/right according to the own view (Ferguson et al., 2015).

about the locations as the object did not always move to another container in the second part of the video. Also, in half of the fillers Jane was present and watched the entire transfer sequence and in the other half she left before the second part of the video as in the experimental trials. Moreover, to prevent participants from knowing whose name was going to be mentioned before they heard 'Jane' during the experimental trials, some of the audio descriptions in the fillers described where someone else would look for the object. In four of the filler trials participants had to infer events according to Jane as in the experimental trials. However, three fillers were according to Sarah's perspective (the fully informed character) and participants would hear "Sarah will look for the [object] in the container on the [left/middle/right]". Additionally, three fillers were based on a stranger's perspective (e.g. "A stranger will look for the [object] in the container on the [left/middle/right]" and six others in reality (e.g. "It's true that the [object] is in the container on the [left/middle/right]"). To ensure participants were paying attention comprehension questions were randomly presented at the end of half of the experimental trials and half of the filler trials.

### 3.2.3 Individual measure tasks

Participants' inhibitory control, working memory and empathy were measured. To measure inhibitory control participants completed a simple and a complex Go-NoGo task as in Chapter 2. Both tasks had Trial Type (Go, NoGo) as a within-participant factor. The Simple Go-NoGo task was based on the standard Go-NoGo paradigm (Simmonds et al., 2008). A coloured square was presented on the left or right side of the screen in each trial. If the square was green (a Go trial) participants had to indicate which side of the screen it appeared on. If the square was red (a NoGo trial) participants had to inhibit their response. In the Complex Go-NoGo task the only difference was that it used yellow and blue squares and included a 1-back working memory (WM) requirement (see Simmonds et al., 2008). The 1-back WM requirement was that participants had to indicate on which side of the screen the square was shown (Go trials) except when a blue square was preceded by a yellow square (NoGo trials).

Working memory was measured using the OSPAN task (La Pointe & Engle, 1990; Turner & Eagle, 1989), and the backward Digit Span subtest of the Wechsler

Adult Intelligence Scale-Third edition (WAIS-III; Wechsler, 1997). In OSPAN task participants were presented with a mathematical equation (e.g.  $(8/2) + 3 = 7$ ) and they had to respond as to whether they thought the answer shown was true by pressing 'y' or false by pressing 'n'. They were then presented with a word on the screen that they have to read aloud and then press the 'spacebar' to proceed. Either 2, 3, 4, or 5 equations / word pairs would be presented in a row after which participants were asked to type in the words they had read out loud in the order they were presented to them. There were 12 trials in total. Responses were recorded using the keyboard and the score was calculated by summing the number of words in correctly recalled word sequences, i.e. where all words were recalled correctly in the right order.

Empathy was assessed using the Interpersonal Reactivity index (IRI; Davis, 1983). The questionnaire contains 4 subscales, i) perspective-taking (e.g. *I sometimes find it difficult to see things from the "other guy's" point of view.*), fantasy (e.g. *I daydream and fantasize, with some regularity, about things that might happen to me.*), empathic concern (e.g. *I often have tender, concerned feelings for people less fortunate than me.*) and personal distress (e.g. *In emergency situations, I feel apprehensive and ill-at-ease.*). There were a total of 28 statements, seven for each subscale. Participants were asked to indicate on a 5-point letter scale (A= *does not describe me well*, to E= *describes me very well*) how well each statement described them. The higher the score indicate higher level of empathy for all subscales.

#### 3.2.4 Procedure

Participants were tested individually and completed the various tasks (False-belief, Go-NoGo, OSPAN, Digit Span, IRI questionnaire) in a single session of approximately 45 minutes. The false belief task was always administered first. Eye-movements during the False-belief task were recorded using a Tobii TX300 eye-tracker at a sampling rate of 300Hz. Stimuli for the False-belief task and OSPAN were presented using E-Prime 2 and the Go-NoGo tasks were programmed in Cogent ([www.vislab.ucl.ac.uk/Cogent/index.html](http://www.vislab.ucl.ac.uk/Cogent/index.html)) running in Matlab 7.0 (MathWorks). The IRI questionnaire was presented through Qualtrics.



For the false-belief task participants were given the following instructions as in Ferguson and colleagues (2015): “In this experiment you will watch short videos, each of which will be followed by a still frame from that video and a spoken description of events. Your task is simply to watch and listen and respond to the comprehension questions when prompted”. After the instructions were given, the experimenter introduced the two characters (Sarah and Jane) by name by showing them a still image of the two characters in the first stage of the video (see **Figure 7**), with the name of each character written by the character’s side. Participants were told that it was very important for them to remember which character is which. To ensure that all participants had a clear understanding of which character was Sarah and which character was Jane they were tested before the start of the experiment by hiding the names from the picture and asking them to point to each character and state their name. Participants were then presented with two filler practice trials (one about what Jane will look for and one about Sarah) to familiarise themselves with the task.

A centrally-located fixation cross was presented at the beginning of each trial, and the trial was initiated when participants successfully fixated it for at least 1s without blinking. The video was then presented which showed the transfer events described above. On average the videos lasted 28s (ranging from 16-53s). Each video was followed by a 500ms blank screen before the final state picture from the video was presented. After a 1000ms preview of the picture the relevant audio target sentence was presented. The picture was presented for a total of 7000ms. The audio typically ended 1-2 s before the trial ended. Trials were separated by a 500ms blank screen. When participants completed the False-belief task they completed the Go-NoGo task, the OSPAN, the backward Digit Span and the IRI questionnaire.

### 3.2.5 Data analysis

#### 3.2.5.1 *Eye-tracking data processing and analysis*

Participant’s eye movements were tracked while the target image (the ‘final state’) was on screen and were processed on a trial-by-trial basis relative to the respective image and sound onsets. For each image, three regions of interest (ROIs) were specified around each container location (left, middle and right). This was done

by mapping spatial coordinates of fixations (in pixels) on each ROI, and if a fixation was located within 20 pixels of a container's perimeter it was coded as a look to that object. Any looks outside of these AOIs were coded as background.

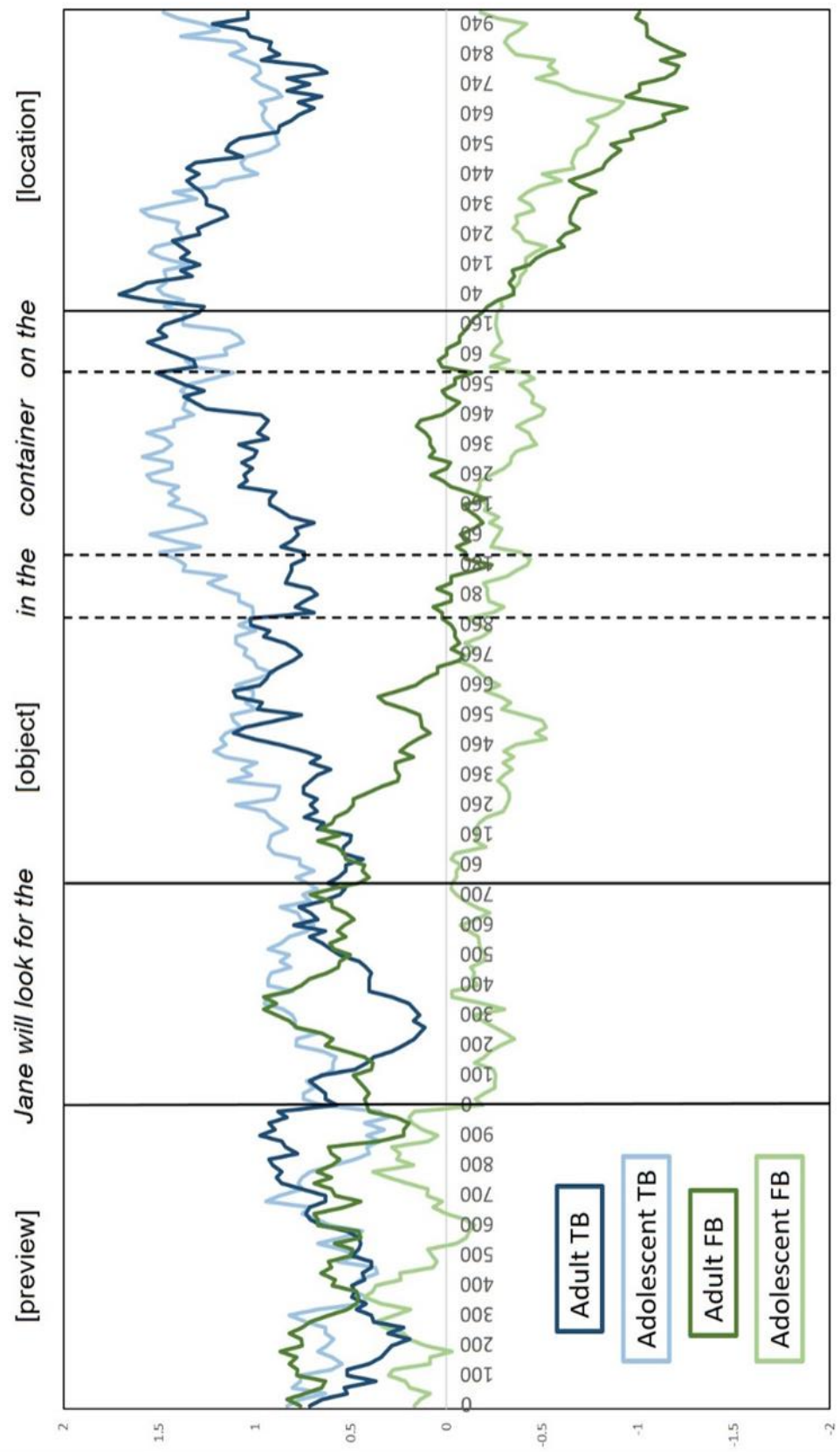
The data was then broken down into 20ms time bins and within each bin fixations were binary coded, with '1' belonging to a ROI (Reality, Belief, Distractor, Background) or '0' if there were no fixations on the ROIs. These fixations were aggregated across participants and items to calculate visual preferences to the *reality location* (the object's final location) and the *belief location* (the initial location) for seven consecutive time windows of interest. The time windows of interest were i) the preview (after image onset but before audio onset), ii) "Jane will look for the", iii) [object], iv) "in the", v) "container", vi) "on the", vii) [location]. **Table 7** shows the average durations for these time windows.

**Table 7** Average time window durations for each condition (timings in ms).

	TB	FB
<i>Pre-object</i>	735	740
<i>[Object]</i>	868	879
<i>"in the"</i>	210	208
<i>"container"</i>	584	585
<i>"on the"</i>	211	220
<i>[Location]</i>	1002	950

For each time window a location-preference score was calculated as in Ferguson and Breheny (2011, 2012) using:  $\log(\text{Reality/Belief}) = \ln (P_{(\text{Reality})} / P_{(\text{Belief})})$ .  $P_{(\text{Reality})}$  refers to the sum of looks to the reality location divided by the total number looks to all ROIs, and  $P_{(\text{Belief})}$  is the sum of looks to the belief location divided by the total number of looks to all ROIs. The output of this calculation is a single value that measures the bias towards each critical location for each condition within each time window. As this measure is symmetrical around zero, positive scores suggest a greater bias to look at the reality location and negative scores indicate a greater bias to look at the belief location. Note that for statistical analysis eye movements were

synchronised to the absolute word onsets and offsets on a trial-by-trial basis. The grand mean for the log-transformed location bias score in each condition and group was plotted to visualise the data (**Figure 8**). Eye-movements here were resynchronized according to individual word onsets (Altmann & Kamide, 2009; Ferguson, Apperly, et al., 2015).



*Figure 8 The average log(Reality/Belief) score for each condition and age group. Note that the dashed and vertical lines indicate absolute onsets and average offsets of words in the target sentence.*

Statistical analyses were carried out using mixed effect regression models for each time window separately: i) [preview], ii) “Jane will look for the”, iii) [object], iv) “in the”, v) “container”, vi) “on the”, vii) [location]. The models were fitted using the ‘lmer’ function in the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) using R (R Core Development Team, 2013). The location-bias score, averaged over a given time window, was used as the dependent variable. Each model included belief (FB vs TB) and age group (Adult vs Adolescent) as fixed effects, deviation coded as FB (-.5) vs TB (.5) and Adolescent (-.5) vs Adult (.5). Models used the maximal random effects structure suggested by Barr, Levy, Scheepers, and Tily (2013), including random effects for participants and items, and crossed random slopes for belief by group on items and a random slope for belief on participants (D. J. Barr, Levy, Scheepers, & Tily, 2013). Random effects were only removed where they lead to non-convergence due to overparameterization. One sample t-tests were used to test whether the fixation bias in each time window was significantly different from zero, which would indicate when participants showed a preference to fixate the reality or belief location in each condition. For all tests a significance level of 5% was used.

#### 3.2.5.2 *Individual measures data analysis*

To examine whether the two age groups differed in any of the individual measures (IC, WM and empathy), one-way ANOVAs were performed. For both the Simple and Complex Go-NoGo tasks, a mean difference in percentage accuracy between NoGo trials and Go trials was calculated for each participant, and median RT was calculated for correct Go trials. For WM a mean score was calculated from the OSPAN task for each participant and the raw score from the backward Digit span were used as dependent variables separately. A sum score from each subset of the IRI questionnaire for each participant was calculated as dependent variables for empathy.

### 3.3 RESULTS

#### 3.3.1 Eye-tracking results

**Table 8** shows the fixed and random effects for the model adopted in each time window, and **Table 9** displays the statistical results from the planned one-sample t-

tests for each condition in each time window. Note that as there were no significant interactions between belief and age group, all post hoc t-tests were performed across all participants (both age groups) for each condition.

**Table 8** Estimates and t-values for each time window of interest.

	Fixed effects				Random effects		
	Intercept	Condition	AgeGroup	Condition:AgeGroup	Participant (Var)	Item (Var)	Residual (Var)
Preview	Estimate (SE) t-Value	0.74 (.41) 1.81	0.11 (.29) 0.39	0.35 (.29) 1.19	0.2 (.52) 0.38	0.19	2.39
Pre-object	Estimate (SE) t-Value	0.53 (.36) 1.48	0.55 (.24) 2.26*	0.58 (.28) 2.08*	-0.8 (.47) -1.72	0.26	1.77
Object	Estimate (SE) t-Value	0.65 (.35) 1.86	1.10 (.28) 3.95***	0.28 (.31) 0.92	-0.69 (.58) -1.13	0.51	1.57
"in the"	Estimate (SE) t-Value	0.41 (.19) 2.15*	1.06 (.27) 3.99***	-0.00 (.23) 0	-0.20 (.51) -0.38	0.18	0.39
"container"	Estimate (SE) t-Value	0.69 (.26) 2.63*	1.78 (.39) 4.58***	0.16 (.34) 0.46	-0.55 (.74) -0.74	0.53	0.73
"on the"	Estimate (SE) t-Value	0.56 (.18) 3.14**	1.63 (.28) 5.78***	0.20 (.30) 0.68	0.34 (.53) 0.63	0.58	0.14
Location	Estimate (SE) t-Value	0.32 (.30) 1.07	2.82 (.29) 9.66***	-0.21 (.32) -0.67	0.37 (.59) 0.63	0.35	1.09

\* p < .05.

\*\* p < .01.

\*\*\* p < .001

**Table 9** One sample t-test results for each condition and time window of interest.

	df	t-Value	p-value
<i>Pre-object</i>			
TB	37	3.98	<0.001***
FB	37	0.73	0.47
<i>[Object]</i>			
TB	37	5.76	<0.001***
FB	37	-0.17	0.87
<i>“in the”</i>			
TB	37	5.71	<0.001***
FB	37	-0.94	0.36
<i>container</i>			
TB	37	7.47	<0.001***
FB	37	-1.00	0.32
<i>“on the”</i>			
TB	37	7.59	<0.001***
FB	37	-1.30	0.20
<i>[Location]</i>			
TB	37	8.76	<0.001***
FB	37	-5.01	<0.001***

\*  $p < .05$

\*\*  $p < .01$

\*\*\*  $p < .001$

Analyses showed no significant effects during the 1000ms preview period, however, during the pre-object window a significant effect of belief was revealed. Post-hoc t-tests revealed that overall, participants were significantly biased to fixate the reality location when they shared Jane’s TB about the object’s location but showed no significant bias to either container when their knowledge of the object’s real location conflicted with Jane’s FB. Interestingly, there was also a significant effect of age group in this pre-object window, showing that adults experienced a stronger egocentric bias to fixate the reality location across true and false belief conditions compared to adolescents. The belief by age group interaction was not significant in this



time window ( $p=.1$ ), however, exploratory analyses revealed that the effect of belief was only significant in the adolescent group,  $Est. = 1.03$ ,  $SE = .36$ ,  $t = 2.83$ ,  $p = .01$ , and not in the adult group,  $Est. = .14$ ,  $SE = .29$ ,  $t = .48$ ,  $p = .64$ .

The significant effect of belief persisted throughout all subsequent time windows. During the object, “in the”, container and “on the” time windows post-hoc analyses revealed that this reflected a significant bias to fixate the reality location on TB trials but no significant bias to either container on FB trials. During the location time windows, post-hoc tests showed that participants were not only significantly biased to fixate the reality location when Jane held a TB about the object’s location but also significantly biased to fixate the belief location when Jane held a FB about the object’s location.

To summarise, the eye-movement data suggest that, overall, the task did not elicit different timings of perspective inference and use between the two age groups. Both adults and adolescents showed significantly different visual biases between TB and FB conditions from the pre-object window. This suggests that an inference about Jane’s belief was made as soon as participants heard whose perspective to take. However, a main effect of age in this pre-object time window suggests that adults’ initial processing is more egocentric than in adolescents. Both groups showed appropriate biases to reality location on TB trials in all time windows of interest. However, although participants seem to infer Jane’s belief from when they hear “Jane”, they do not seem to use this perspective information until the location time window, where they begin to correctly anticipate reference to the belief location on FB trials.

#### *3.3.1.1 Individual measures results*

Results for one way ANOVAs of all individual measures are shown in **Table 10**. Adults and adolescents did not differ in any of the IC and WM measures (all  $ps > .11$ ). They also did not differ in empathy measures (all  $ps > .16$ ), apart from the perspective-taking subscale ( $p = .01$ ).

**Table 10** Means and standard deviations of all individual measure scores and their respective one way ANOVA results.

	Mean (SD)		ANOVA	
	Adult	Adolescent	df	F
<i>Inhibitory control</i>				
<i>Accuracy</i>				
Simple Go NoGo	-0.05 (.07)	-.0.8 (.13)	1, 36	0.44
Complex Go NoGo	-0.23 (.16)	-0.25 (.14)	1, 36	0.19
<i>RT</i>				
Simple Go NoGo	381.94 (23.19)	384.81 (45.59)	1, 36	0.06
Complex Go NoGo	421.62 (46.68)	433.50 (53.64)	1, 36	0.52
<i>Working Memory</i>				
Digit Span	7.65 (2.34)	6.91 (2.02)	1, 36	1.10
OSPAN task	32.12 (5.57)	26.76 (12.31)	1, 36	2.75
<i>Empathy</i>				
Perspective-taking	20.65 (4.15)	17.38 (3.26)	1, 36	7.38**
Fantasy	17.06 (3.52)	18.14 (4.67)	1, 36	0.63
Empathic concern	21.82 (3.38)	19.86 (4.73)	1, 36	2.08
Personal distress	11.82 (4.75)	12.48 (5.72)	1, 36	0.14

\*  $p < .05$

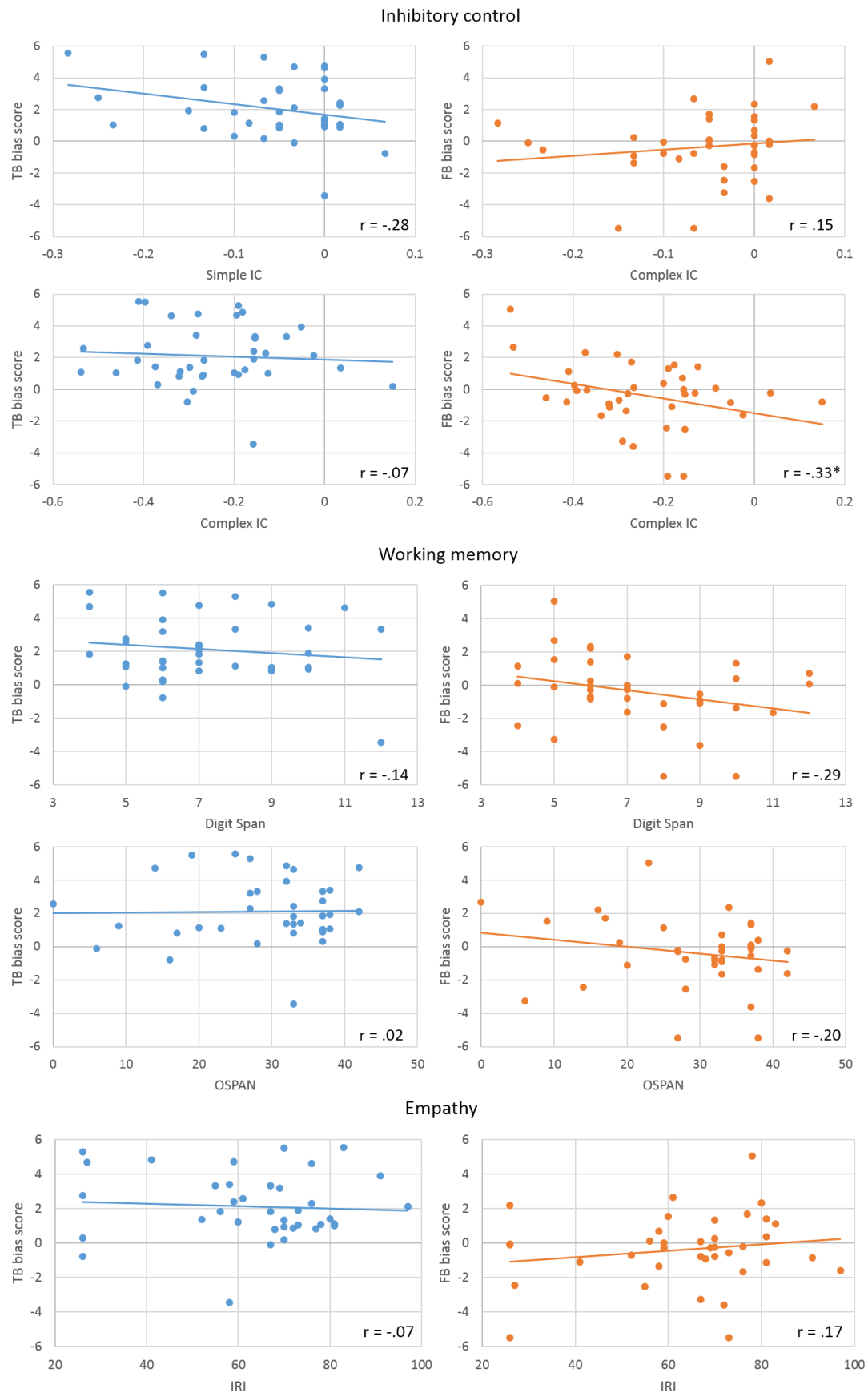
\*\*  $p < .01$

To further explore whether individual differences in IC, WM and empathy had an effect on looks to the correct target, Pearson correlations were performed on participants' IC (simple and complex), WM and empathy scores. An average bias score for each condition over the "container on the" time windows was calculated. Correlations were conducted with this averaged bias score of the two time windows as it is the period that immediately precedes the disambiguating information and therefore, should show the clearest effects of anticipation. The data is presented in **Figure 9**. Results showed that there were no significant correlations for empathy, WM

and the simple IC (all  $p$ 's > .08).<sup>5</sup> Interestingly, complex IC correlated with the bias score in the FB condition [ $r(38) = -.33, p = .043$ ] but not with the bias score in the TB condition [ $r(38) = -.072, p = .669$ ]. This suggests that appropriate biases toward the initial location in the FB condition became more likely and stronger with better performance in the complex IC task.

---

<sup>5</sup> Note that one participant was removed from the simple IC and TB and FB bias score correlation analyses because their score was highlighted as an outlier.



**Figure 9** Correlation between participants bias score (y-axis) in the TB (blue) and FB (orange) condition in the ‘container on the’ region and each individual measure (x-axis).

### 3.4 DISCUSSION

The present study had three main objectives i) to replicate the adult findings from Ferguson and colleagues (2015), ii) to investigate whether adolescents can infer other peoples' actions based on (false) beliefs as early and as spontaneously as adults when they have no explicit reason to do so, and iii) to explore the role of age-related differences in inhibitory control, working memory or empathy in constraining performance during online false-belief reasoning. Adolescent and adult participants' eye gaze was tracked while they were presented with videos that portrayed true belief scenarios where the participant held the same belief as the character, and false belief scenarios where the character's belief was different from their own. This allowed us to track and compare the time course of false-belief reasoning online between adults and adolescents. A difference in looking pattern between TB and FB conditions was examined at seven key points during each trial. Participants' individual measures that have been found to be associated with online ToM use, such as inhibitory control, working memory and empathy, were also measured to check for baseline differences between groups.

Results from the adult participants in the false-belief task replicated Ferguson and colleagues (2015), namely that participants showed significantly different visual biases between TB and FB conditions from the pre-object window, suggesting an inference about Jane's belief was made as soon as participants heard whose perspective to take. As in Ferguson and colleagues (2015) adult participants showed appropriate anticipatory looks to the reality location in TB trials. For FB trials they only made appropriate biases to the initial location until the location was auditorily available. Crucially, adolescents' data showed the exact same pattern as adults. Overall, the two age groups' gaze bias developed in the same way over time, with the adolescents showing a sensitivity to Jane's belief as early as adults did and appropriately anticipating the reality container in TB trials. Both adults and adolescents then seem to be sensitive to other people's perspectives spontaneously even when they have no reason to do so. Interestingly, a main effect of group during the pre-object window suggested that adults experienced a stronger egocentric bias initially than adolescents. In other words, adolescents suffer less egocentric bias, indicating

that not only are adolescents' performance comparable to adults here, but perhaps may actually be better.

The results in this study therefore contrast with previous research that found age-related differences in young adolescents during a false-belief task (Devine & Hughes, 2013). The results are also not in line with results found in Chapter 2 where both younger and older adolescents showed age-related differences from adults. Rather than conclude that we simply failed to find an effect where there should be one, I propose that this finding is in line with our prediction that, at least in comparison to the DT, adolescents' performance in the FB task would be comparable with the adult group as the factors I suggested constrain adolescent performance in the DT task are not present in the current task. Firstly, the mental state information in the DT is not as prominent as in the FB task, as the available cues in the DT are less salient. The cues in the FB task are more available as participants see Sarah moving the object. Moreover, participants observe Jane leaving the room which makes it more salient that she cannot see where the object will move next. By comparison, participants in the DT have to infer mental state information from the instruction given at the beginning of the experiment and a relatively small amount of visual information about the perspective of the director (being the opaque background) in each trial.

Secondly, the DT involved more complex integration of ToM and EF processes as participants use mental state information from less available cues while inhibiting linguistic cues in order to infer someone else's belief. In addition, the participant has to use the information that is based on someone else's incomplete belief to plan their own action and execute it very quickly.

As discussed in Chapter 2, models of perspective-taking found in Hanna et al. (2003) and Brown-Schmidt et al. (2008) suggest that perspective information can be integrated from the outset. On this view, the extent to which mental perspective information is integrated on-line depends on the extent to which other constraints are conflicting and how salient or available the perspective information is (Brown-Schmidt & Hanna, 2011). Moreover, it depends on how sensitive a participant is to such information (Brown-Schmidt, 2012) or their ability to integrate that information with linguistic and visual contextual information (Hanna et al., 2003). Therefore, I suggest

then that where adolescents' performance differs from adults on the DT is not necessarily due to a lesser ability to infer other people's belief/perspective, but it may be due to less efficiency in exploiting subtle cues and integrating that information online. This is a hypothesis that requires further investigation.

An alternative reason why age differences were not found here comes from evidence in Ferguson and colleagues (2015). They have shown that in an active version of this false-belief task adult participants can be much faster in anticipating not only the reality container in TB trials but also the initial container in FB trials. While in the passive task participants were asked to 'look and listen', in the active version of the task participants were asked to press one of three keys that corresponded to the three containers in order to select the container that would complete the sentence correctly. Adults in the active group showed appropriate anticipatory looks to the initial location during FB trials as soon as they heard whose perspective to take ("Jane" onwards). The authors argue that belief inferences are made earlier and the information is used to generate expectations sooner when participants are actively engaged in the task in comparison to when participants are passive observers. Perhaps then, the passive task was not engaging enough and therefore, age-related differences could not be disentangled. In both the Silent films and the Director task, the participants had an explicit reason to keep track of the director's/character's perspective. Further research is required to explore this possibility, perhaps by testing adolescents with the active task in order to explore whether the fact that they have to respond explicitly delays their performance in comparison to adults.

As mentioned above, the results here are not in line with results from Devine & Hughes (2013) who found age-related differences in young adolescents during a false-belief task. However, it is difficult to compare the Silent films task and the FB task as the former involves a production component. Participants in the Silent films task are required to explain what happened in the videos verbally and are scored on that basis. This involves not only making the right inferences when watching the film, but also putting that into words. This extra level of production demand may have impacted on their performance. Having said that, it may be that the video scenarios are in general

more complex than in the FB task as the Silent films task includes a variety of scenario types (false belief/ignorance/deception etc.).

Consistent with the results from Chapter 2 is the fact that both age groups infer Jane's perspective spontaneously as soon as they hear whose perspective to take. In the DT results showed that when participants (both adolescents and adults) responded correctly their gaze data showed an immediate response when the critical verbal input is presented. Eye-tracking data showed early eye gaze differences between trials where the director's perspective was taken into account versus when it was not. In correct responses when participants engaged perspective-taking, all age groups display the same pattern of gaze bias over time. In other words, both adults' and adolescents' gaze data suggests that both groups commit to a target at the earliest point in the linguistic stimulus, whether that was the correct target or not. However, for reasons discussed above, I propose that adolescents did not engage in perspective-taking as efficiently as the adult group during the DT due to the higher demands of integrating ToM online from subtle cues, which given the current results seems to be a constraining factor for adolescents' performance (Chapter 2; Symeonidou et al., 2016).

Despite the early divergence of eye-movements by condition, both age groups showed a delay in predicting the appropriate container during FB trials, in comparison to TB trials where they showed clear anticipatory looks to the appropriate container. According to the Perspective-Adjustment model (Epley et al., 2004; Keysar et al., 2000, 2003) (or reality bias as per Birch & Bloom, 2007), this delay would be explained a default initial bias to take an egocentric perspective which is later adjusted, during a second process to the speaker's perspective. However, this does not explain why listeners show significantly different visual biases between TB and FB trials from when they hear whose perspective to take, showing a clear early sensitivity to the different perspectives in both conditions. If egocentric biases are the default in ToM processing, then a difference in visual biases should not be occurring early on in the trial. Instead divergence would occur at later stage when listeners would adjust their perspective according to the character's knowledge. Furthermore, results from the active group in Ferguson and colleagues (2015) do not show an initial egocentric bias as participants correctly predict the initial location in FB trials from when they hear whose perspective



to adopt. Added to this, results from Chapter 2 are also inconsistent with the Perspective-Adjustment model.

Ferguson and colleagues (2015) provide an alternative explanation for the delayed FB predictions in the passive group of their study. In both current results and in Ferguson and colleagues (2015) participants do not show an initial preference to either location. According to Ferguson and colleagues (2015) this means they are holding multiple representations of the same object for each perspective appropriate location during the ambiguous period (Altmann & Kamide, 2009). In other words, in both conditions (active and passive) participants held two competing representations - an egocentric and a perspective-appropriate. In the case of the active group however, the perspective-appropriate representation is actively maintained due to the explicit nature of the task. It is suggested, that the delayed bias to the initial location in the passive compared to the active group then is because the alternative perspective is not as actively maintained which results in a bigger influence from the egocentric perspective.

As with the eye-movement data, we did not find any global differences in IC and WM, suggesting that our age groups were well matched on these skills. Interestingly, complex IC scores correlated with the bias score in the FB condition (i.e. appropriate biases toward the initial location became more likely and stronger with better performance in the complex IC task), but not with the bias score in the TB condition. This is not surprising as participants need to inhibit their own knowledge about reality to infer the character's belief. Adolescents and adults also matched in empathy measures with the exception of the perspective-taking subscale on the IRI questionnaire, where the adolescent group scored lower than the adult group. However, note that the perspective-taking being tested on the empathy questionnaire is not the same perspective-taking process discussed in false-belief reasoning. Davis (1983) describes the perspective-taking subscale as "the tendency to spontaneously adopt the psychological point of view of others" (Davis, 1983, pp. 113–114). The items on the PT subscale do not tap into whether someone is sensitive to reason about someone else's false-belief. Rather, the questions focus on emotional components such as imagining how others would feel if they were in the same situation (e.g. *Before*

*criticizing somebody, I try to imagine how I would feel if I were in their place or When I'm upset at someone, I usually try to "put myself in his shoes" for a while).* Perhaps this is an indication that one possible constraint for adolescents in ToM processing is when situations have an emotional component to consider. Indeed, this would be in line with the results from Sebastian and colleagues (2012) that found age-related differences between adolescents and adults in affective ToM cartoon scenarios that involved emotional aspects and not with the cartoons that involved cognitive ToM. This hypothesis requires further investigation.

In conclusion, the current study revealed that adolescents consider other people's perspective as early and as spontaneously as adults even when they have no explicit reason to do so. In fact, adults might suffer from a stronger ego-centric bias during initial processing. Crucially, the current results suggest that attending to and integrating relevant ToM information rapidly from subtle cues is a possible constraining factor for adolescents in online ToM processing. As such, the results provide online evidence that adolescents can spontaneously infer other people's perspective during false-belief reasoning when available cues are salient, and EF demands are lower. Further research is required to disentangle what other factors constrain adolescents' ToM processing. One possible direction is investigating their sensitivity to available cues to mental state information when emotional aspects are involved and ToM demands are higher but EF demands remain low. Specifically, a task where higher-order ToM is involved but that does not involve participants planning their own action or response based on someone else's perspective and there are no conflicts from linguistic cues or knowledge from the self-perspective.

## 4 CHAPTER 4: SECRETS AND ADOLESCENTS

---

### 4.1 INTRODUCTION

In Chapter 2 I investigated the development of ToM use through eye tracking in young adolescents (9-13 years old), older adolescents (14-17.9 years old), and adults (19-29 years old), using a computerised version of the Director task. Results provided evidence for prolonged development of online use of ToM in adolescents and while we found that inhibitory control might be a factor in the poorer performance, results suggested that other factors were at play. Given that the timecourse of gaze bias formation for correct trials did not differ between adults and younger age groups, we proposed other factors that had not to do with the ability to make ToM inferences per se, but the ability to integrate information relevant to the speaker's beliefs and referential intentions rapidly in on-line reference assignment. Given these results and the challenges of the director task in terms of ToM integration and inhibitory control, I conducted a new eye-tracking study in Chapter 3 to investigate whether adolescents can reliably infer others' (false) beliefs as spontaneously and early as adults, even without being given an explicit instruction to do so. This paradigm allowed us to reduce the demands of the director task by recording implicit perspective taking (i.e. no task), providing more salient perspective cues, and eliminating linguistic conflict. Here, all participants showed significantly different visual biases between TB and FB conditions as soon as they heard whose perspective to take (Jane's). Moreover, participants showed appropriate anticipatory looks to the reality location in TB trials, but in FB trials they only made appropriate biases to the initial location right when they heard the disambiguating information. This suggests that when task demands are lowered adolescents are sensitive to other people's perspective as early and as spontaneously as adults even when they have no explicit reason to do so. In fact, results suggested that adults might suffer a stronger initial egocentric bias than adolescents, which is further evidence of adolescents not performing more poorly from adults.

In summary, in two different kinds of task, we have two different kinds of outcome – a difference between adolescents and adults in the DT, and no difference in

the FB task. I attributed these differences to factors either related to EF demands of the different tasks or to a difference relating to processing efficiency – an ability to detect and integrate cues to mental state information. I have argued that the two tasks are essentially the same in terms of the inferential complexity involved; in both cases, an inference has to be made based on a first-order belief of another agent, which differs from the participant’s belief. Therefore, the question remains as to whether there might be a difference between adolescents and adults in a task that involves a greater degree of ToM inferential complexity. A task with greater ToM inferential complexity may require participants to predict or explain another person’s behaviour based on higher-order mental states – for example, beliefs about beliefs, desires about what other people believe, and so forth.

In this chapter, I aim to address this issue by testing adolescents on a task that is arguably comparable to the FB task in terms of EF demands and salience of relevant mental state information, but differs in terms of the complexity of the ToM inference. Specifically, this current study will investigate whether adolescents can use knowledge about a character’s preferences and higher-order desires to make complex ToM inferences and predict that character’s subsequent behaviour as quickly as adults.

#### 4.1.1 The current study

In the current experiment we adopted a visual world paradigm from Ferguson and Breheny (2011) where participants were presented with two-sentence stories. Sentence (1) introduced a property of the character such as a personal preference and a context in which that character is either happy for others to know that property (1a) or the character does not want other people to know about that property (1b). Therefore, in the *open* condition (1a) the basic preference and higher-order desires match, whereas in the *secret* condition (1b) the character’s basic preferences and higher-order desires are in conflict.

1. a. Helen doesn’t care who knows that she dislikes vegetables.  
b. Helen is very secretive about the fact that she dislikes vegetables.

Sentence (2) described the character performing an action that is consistent with the context in sentence (1). For example, in a context such as (1a) Helen would behave in accordance with her personal preferences and appropriately eat something other than vegetables such as meat (2a). In contrast, in a context such as (1b) Helen would adapt her behaviour to fit with her desire to keep personal preferences a secret so would eat vegetables (2b). Participants heard sentence (2) while a visual display was presented which included images that were consistent with the character's two possible choices of action. Although one can predict the appropriate action based on the contextual information given in (1), Ferguson & Breheny (2011) argue that both conditions involve higher-order ToM reasoning as the character's action is not only based on their basic preference but their desires and intentions on how they want to be viewed by others. Moreover, they observe that the *secret* condition may be more demanding as the character's basic preference (Helen's dislike of vegetables) and their high-order intention of keeping this preference a secret, are in conflict. This means that as both the basic preference and higher-order desires of the character as equally salient, when predicting the character's action, a participant has to ignore any predictions that are based on the character's basic preference and they have to make both inferences in order to make a correct prediction. While in the *open* condition, basic preferences and intentions point to the same target, in the *secret* condition they do not.

2.     a. When Helen goes to dinner parties she makes a show of eating meat.  
       b. When Helen goes to dinner parties she makes a show of eating vegetables.

The results from adult participants in Ferguson & Breheny (2011) showed, in both conditions, that participants employed different visual biases very early on in the linguistic stimulus, despite the fact that the character's basic preference was the same in both contexts. That is, participants exploited information about both their basic preference and higher-order desires in both conditions. In the open condition where there was no conflict between the character's basic preferences and high-order desires, adults reliably anticipated the target from when they heard the character's basic goal

(e.g. *going to dinner*). Similarly, in the secret condition when the character's basic preference and higher-order desires were in conflict, participants anticipated the secret-appropriate target. In both conditions then participants successfully used ToM reasoning to generate expectations about the character's subsequent behaviour. However, the bias in the secret condition was delayed in comparison to the open condition as participants only showed a preference to the target from the post-ambiguous noun region (the offset of 'dinner parties' and onset of 'makes a show'), suggesting that character's basic preferences influenced expectations.

The 'secrets' paradigm is well suited to address the issues raised above in the following ways. Firstly, based on previous results (Ferguson & Breheny, 2011), adults are successful in this paradigm as they are able to generate expectations using high-order ToM reasoning rapidly and this will allow us to examine whether adolescents' performance is comparable to adults. Secondly, this paradigm eliminates two possible factors that have been argued as possible constraints during ToM processing that are typically found in perspective taking tasks, including the DT. The first is it avoids low-level linguistic interference. Most false-belief/perspective taking tasks involve an ambiguous noun or expression creating a conflict from the linguistic input. However, it has been shown that performance on a perspective-taking task can be affected by the nature of the verbal stimulus (Barr, 2008). The second constraint is the 'curse of knowledge' (Birch & Bloom, 2007) which is the idea that during such tasks a participant always has a more informed perspective than the other person, i.e. there is conflicting knowledge from the self-perspective. The secrets paradigm involves stories that require participants to reason about others' behaviour based on shared knowledge about 'reality' and personal preferences, without conflict from the linguistic input or knowledge from the self-perspective. Finally, as the secrets paradigm provides a task that reduces demands on IC and other executive mechanisms that might be costly but that involves more complex, higher-order ToM inferences, we can explore the development of ToM abilities from a different perspective to the Director Task.

As in Chapter 3, to control for individual differences, participants' inhibitory control (IC), working memory (WM) and empathy were measured due to previous associations that have been found with ToM (see Chapters 1, 2 and 3).

To sum up, this study aims to i) replicate the adult results from Ferguson & Breheny (2011), ii) to investigate whether younger and older adolescents can use knowledge about a character's basic preferences and higher order desires, even when they are in conflict, to make complex ToM inferences and predict that character's subsequent behaviours as quickly as adults. In the current study we predict similar results as in Ferguson & Breheny (2012) for the adult data, namely there should be a significant difference in looks between the open and secret conditions early on from when they learn the character's goal/action. Moreover, participants should show anticipatory looks to the appropriate target before they hear the disambiguating information, with perhaps a minor delay in the secret condition relative to the open condition. For the adolescent data, if there is ongoing development in the ability to perform more complex, higher-order ToM inferences, then we should see a delay in the time-course of target bias formation in both the open and secret conditions, in comparison to adults. Moreover, given previous results (Chapter 2, Symeonidou et al., 2016; Dumontheil et al., 2010), we predict that the younger adolescents will show a delay from the older adolescents for both conditions.

## **4.2 METHOD**

### **4.2.1 Participants**

Fifty-two participants took part in this study, of which 17 were adults (24-36 years old,  $M = 27.32$ ,  $SD = 3.57$ ), 18 were older adolescents (14-18 years old,  $M = 16.70$ ,  $SD = 1.39$ ), and 17 were younger adolescents (9-13.9 years old,  $M = 11.81$ ,  $SD = 1.43$ ). All participants were native English speakers. Adult participants were recruited from the UCL Psychology participant pool while adolescents were mostly recruited from London schools. Parents/guardians of all adolescent participants as well as all adult participants were given information sheets prior to the study and informed consent was obtained. This study was approved by the UCL Research Ethics Committee.

#### 4.2.2 Stimuli and design

##### 4.2.2.1 *Secrets task*

The task design and stimuli were based on a previous study by Ferguson & Breheny (2011). The experiment was a 2 x 2 mixed design, with Context (Open vs false belief) as the within-participant factor and Age group (Adult, Older Adolescent, Younger Adolescent) as the between-participants factor. There were 16 sets of experimental pictures with their respective auditory description in one of two conditions. Most of the visual stimuli that were used in this study were the same as in Ferguson & Breheny (2011) with the exception of a few that were not appropriate for the younger adolescent group (e.g. containing alcohol). Those stimuli were replaced with age-appropriate stimuli, as in the original study. Due to the few changes in the visual stimuli all auditory stimuli were recorded again from a female native British English speaker so that all auditory stimuli were from a single recording session for all trials.

**Table 11** shows an example of the experimental auditory stimuli and **Figure 10** shows their respective visual display. Each visual stimulus contained four images: Character, Open Referent (pink car), Secret Referent (green car), and a Distractor (weights) which was an irrelevant object. The position of the four different types of pictures differed across items to prevent participants from creating viewing strategies. The auditory stimulus for each item consisted of two sentences. In the first sentence a fact about the story character (e.g. *Tom's favourite colour is pink*) was introduced within an open context ("Tom is always telling people that...") or a secret context ("Tom doesn't want anyone to know that..."). Sentence two was a description of an event (e.g. *buying a new car*) that referred to an open relevant referent ("...he deliberately chose a *pink* car") or a secret relevant referent ("...he deliberately chose a *green* car"). All character's actions described in the second sentence were consistent with their personal preference.



**Table 11** Examples of experimental sentences.

---

*Open*

Tom is always telling people that his favourite colour is pink

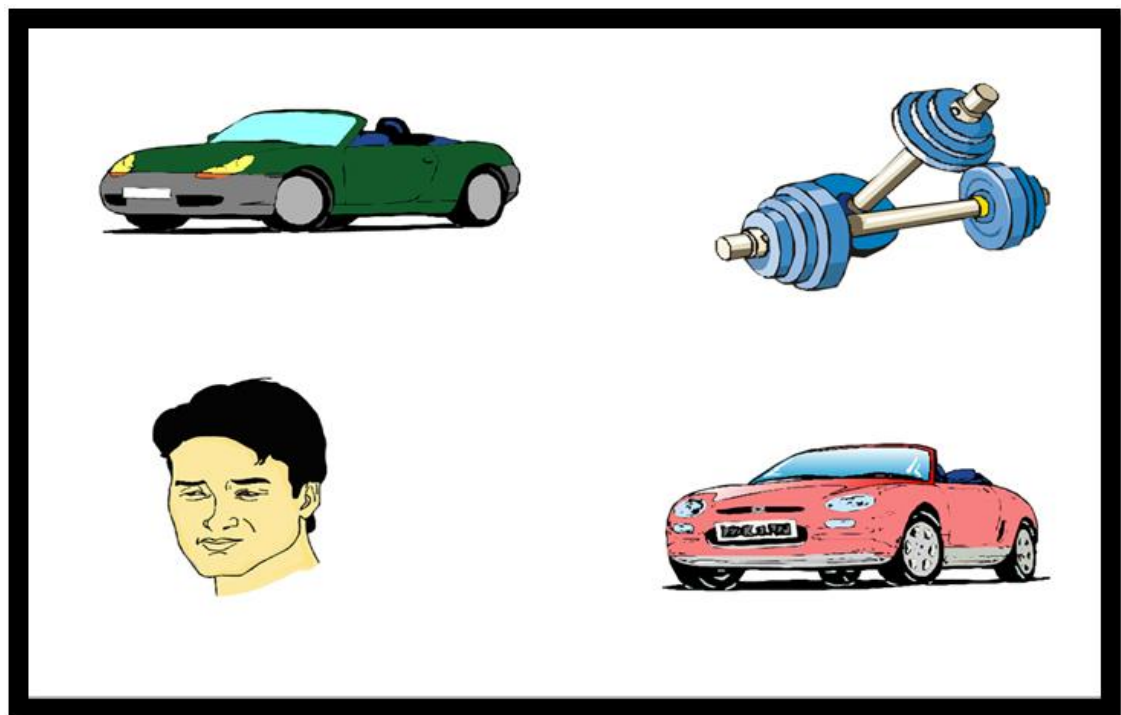
Last week Tom bought a new car and he deliberately chose a pink car.

*Secret*

Tom does not want anyone to know that his favourite colour is pink.

Last week Tom bought a new car and he deliberately chose a green car.

---



**Figure 10** Example of visual stimulus that participants viewed while they heard the target sentence (sentence 2).

One version of each item was assigned to one of two lists and participants were randomly assigned to one of these lists. This ensured that each participant only heard each target sentence once in one of the two conditions. Each list contained 16 unique experimental items (see **Appendix III**), eight in each condition (open and secret context). Sixteen filler trials were also included in each list which were interspersed randomly among the 16 experimental trials. The fillers consisted of similar visual stimuli, paired

with two auditory sentences which required participants to make linguistic inferences (e.g. “Because Greenland has really cold winters, it is popular with tourists of all ages. We can go to Greenland in November to relax with family.”). Crucially, however, they did not involve ToM reasoning, surprising events or unexpected endings.

Comprehension questions were included after half of the experimental trials and half of the fillers. The questions were either about the visual or auditory stimuli and participants had to respond with a ‘yes’ or ‘no’ answer. No feedback was given to the participants for their responses to the questions.

#### 4.2.2.2 *Individual difference measures*

Participants’ IC, WM and empathy were measured. As in Chapters 2 and 3, participants completed a simple and a complex Go-NoGo task to measure IC. Both tasks had Trial Type (Go, NoGo) as a within-participant factor. The Simple Go-NoGo task was based on the standard Go-NoGo paradigm (Simmonds et al., 2008). A coloured square was presented on the left or right side of the screen in each trial. If the square was green (a Go trial) participants had to indicate which side of the screen it appeared on. If the square was red (a NoGo trial) participants had to inhibit their response. In the Complex Go-NoGo task the only difference was that it used yellow and blue squares and included a 1-back working memory (WM) requirement (see Simmonds et al., 2008). The 1-back WM requirement was that participants had to indicate on which side of the screen the square was shown (Go trials) except when a blue square was preceded by a yellow square (NoGo trials).

WM was measured using the backward Digit Span subtest of the Wechsler Adult Intelligence Scale-Third edition (WAIS-III, Wechsler, 1997) for all participants. The older adolescent and adult groups also completed the OSPAN task (La Pointe & Engle, 1990; Turner & Eagle, 1989). However, this task was too difficult for the young adolescent group therefore, it was not included in the analysis.

Empathy was assessed using the Interpersonal Reactivity index (IRI; Davis, 1983) for the older adolescent and adult group. The questionnaire contains 4 subscales, i) perspective taking (e.g. *I sometimes find it difficult to see things from the “other guy’s” point of view.*), fantasy (e.g. *I daydream and fantasize, with some regularity, about things*

that might happen to me.), empathic concern (e.g. *I often have tender, concerned feelings for people less fortunate than me.*) and personal distress (e.g. *In emergency situations, I feel apprehensive and ill-at-ease.*). There were a total of 28 statements, seven for each subscale. Participants were asked to indicate on a 5-point letter scale (A= *does not describe me well*, to E= *describes me very well*) how well each statement described them. The higher the score indicates higher level of empathy for all subscales.

As the IRI was too difficult for the younger adolescent group, their empathy was measured using the Empathy Questionnaire for Children and Adolescents (EmQue-CA; Rieffe, Ketelaar, & Wiefferink, 2010) which is aimed at adolescents between the ages of 9-16. The EmQue-CA includes 3 subscales: Affective empathy (e.g. *If my mother is happy, I also feel happy*), Cognitive empathy (e.g. *When a friend is angry, I tend to know why*), and Prosocial Motivation (e.g. *If a friend is sad, I like to comfort him*). There are a total of 18 statements, nine for each subscale. Participants were asked to mark on a three-point scale whether the sentence was *often true*, *sometimes true* or *not true* for them. A higher score indicates a higher level of empathy. Note that the adolescents that were 12 years or older completed both the EmQue-CA and the IRI questionnaires (with the exception of two older adolescents who did not fill in the EmQue). This was done to ensure that the empathy scores from the two different questionnaires correlated which they did [ $r(23) = .64, p = .001$ ]. However, the main analysis for the older adolescent group was conducted with the IRI scores as the EmQue-CA questionnaire was not an appropriate level of difficulty for participants over the age of 16 in the older adolescent group.

#### 4.2.3 Procedure

Participants were tested individually and completed the various tasks (Secrets, Go-NoGo, OSPAN, Digit Span, IRI/EmQue-CA questionnaire) in one session of approximately 45 minutes. Eye-movements during the Secrets task were recorded using a Tobii TX300 eye-tracker at a sampling rate of 300Hz. Stimuli for the Secrets task and OSPAN were presented using E-Prime 2 and the Go-NoGo tasks were programmed in Cogent ([www.vislab.ucl.ac.uk/Cogent/index.html](http://www.vislab.ucl.ac.uk/Cogent/index.html)) running in Matlab 7.0 (MathWorks). The IRI and EmQue-CA questionnaires were presented on a computer through Qualtrics.

For the secrets task participants were given the following instructions as in Ferguson & Breheny (2011): “In this experiment you will hear short spoken passages and during the second sentence a picture will also be displayed. We are interested in how the pictures help you understand the spoken language.”

A centrally-located fixation was presented at the beginning of each trial that would only move on to the trial if participants successfully fixated on it for at least 1s without blinking. This ensured that the eye-tracker was still tracking the participant’s eye-gaze and the participant was focusing his/her gaze at the centre of the screen before each trial begun. After a successful fixation, the fixation cross remained on the screen while participants heard the first sentence. They were told to keep looking at the fixation cross while they heard Sentence 1. Participants were then saw a 100ms blank screen after which the target picture was presented. After a 1000ms picture preview participants the audio for Sentence 2 was initiated. The picture remained on screen for 9s and Sentence 2 ended approximately 1-2s before the end of the trial. When participants completed the Secrets task they were given the Go-NoGo task, the OSPAN, the backward Digit Span and the IRI and EmQue-CA questionnaires. Depending on the age of the participant not all participants completed all of the above tests as explained above.

#### 4.2.4 Data analysis

##### 4.2.4.1 *Eye-tracking data processing and analysis*

Participant’s eye movements were tracked while Sentence 2 (the target sentence) was presented and the picture was on screen. For each image, the regions of interest (ROIs) were specified around each smaller image on the picture (character, open referent, secret referent and distractor). This was done by mapping spatial coordinates of looks (in pixels) onto the ROIs and if a look was located within 20 pixels around an image’s perimeter it was coded as a look to that object. Any looks outside of the 20 pixels were coded as background.

The data was broken down to 20ms bins and within each bin looks were coded as ‘1’ belonging to a ROI (Character, Open, Secret, Distractor, Background) and ‘0’ if

there were no looks in any of the above ROIs. These looks were aggregated across participants and items to calculate a probability of looks to the *open referent* and the *secret referent* for five different time windows: i) ambiguous noun (e.g. “car”), ii) post-ambiguous noun, iii) adverb (e.g. “deliberately”), iv) transitive verb, and v) disambiguating noun. **Table 12** shows the average durations for these time windows.

**Table 12** Average time window durations for each condition (timings in ms).

	Open	Secret
Ambiguous noun	534	541
<i>Post-ambiguous noun</i>	480	502
<i>Adverb</i>	825	836
<i>Transitive verb</i>	597	612
<i>Disambiguating noun</i>	1109	1081

For each time-window a referent preference score was calculated as in Ferguson and Breheny (2011, 2012) using:  $\log(\text{Open/Secret}) = \ln(P_{(\text{Open})} / P_{(\text{Secret})})$ .  $P_{(\text{Open})}$  refers to the sum of looks to the open referent divided by the total number looks to all ROIs within that trial and  $P_{(\text{Secret})}$  is the sum of looks to the secret referent divided by the total number of looks to all ROIs within that trial. The output of this calculation is a single value that measures the bias towards each critical referent for each condition within each time window. As this measure is symmetrical around zero, positive scores suggest a higher proportion of looks to the open referent and negative scores indicate a higher proportion of looks to the secret referent. Note that for statistical analysis eye-movements were synchronised to the absolute word onsets and offsets on a trial-by-trial basis. The grand mean for the log-transformed referent bias score was plotted to visualise the data (**Figure 11**). Eye-movements here were resynchronized according to individual word onsets (Altmann & Kamide, 2009; Ferguson et al., 2015).

Statistical analyses were carried out with mixed effect regression models for each time window separately: i) ambiguous noun (e.g. “car”), ii) post-ambiguous noun,

iii) adverb (e.g. “deliberately”), iv) transitive verb, and v) disambiguating noun.<sup>6</sup> The models were fitted using the ‘lmer’ function in the lme4 package (Bates et al., 2015) using R (R Development Core Team, 2013). The referent-bias score was used as the dependent variable which, as described above, is participant’s bias to look at the open/secret location. Each model included context (Open vs. Secret) as a fixed effect and it was deviation coded as Secret (-.5) vs Open (.5). Age group was also a fixed effect in each model. However, due to the fact that it contains three levels, two deviation coded contrast schemes were applied to the Age group variable to fit within the mixed-effect model analyses: Contrast 1 = Adult (1), Older Adolescent (-.5), Younger Adolescent (-.5); Contrast 2 = Adult (0), Older Adolescent (.5), Younger Adolescent (-.5). This contrast coding allowed us to compare the adult group with both adolescent groups together (Contrast 1), and compare the older adolescent group with the younger adolescent group without the adult group (Contrast 2). Where post-hoc analyses were required to follow up on significant interactions, models were re-levelled to include the condition of interest as the reference level. One sample t-tests were also used to follow up on significant effects of context to test whether gaze bias was significantly different from zero which would indicate a bias to the open referent or the secret referent location. The models had a maximal random effects structure with random effects for participants and items, and crossed random slopes for context by group on items and a random slope for context on participants. Where the model did not converge due to over-parameterization, random effects were removed. For all tests a significance level of 5% was used.

#### 4.2.4.2 *Individual measures data analysis*

To examine whether there were age group differences in any of the individual measures (IC, WM and empathy) one-way ANOVAs were performed between the three groups. For both the Simple and Complex Go-NoGo tasks, a mean difference in percentage accuracy between NoGo trials and Go trials as a measure of accuracy was

---

<sup>6</sup> Note that the preview window is not included in the current study as it has been shown before in Ferguson and Breheny (2011) that prior to linguistic input participants show an initial visual preference for the open referent (*pink car*) compared to the secret referent (*green car*). This is unsurprising and perhaps expected since the open referent is explicitly mentioned in Sentence 1 in both conditions.

calculated for each participant. Median RT was calculated for correct Go trials. One-way ANOVAs were performed to examine the effect of Age group on participants' accuracy and RTs in each task. For WM the raw score from the backward Digit span of each participant was used as dependent variables separately. For empathy a sum score was calculated for each participant from the EmQue-CA for the younger adolescent group and from the IRI for the older adolescent and adult group. One-way ANOVAs were performed using the EmQue-CA scores to examine age group differences between the younger and older adolescent groups and the IRI scores were used to compare the older adolescent group with adults. Notably, given the nature of the questionnaires, we could not directly compare the young adolescent group and adults in this individual measure. To explore whether individual differences in IC, WM or empathy had an effect on performance in the secrets task, Pearson correlations were performed between participants scores on each individual measure (IC, WM and empathy as explained above) and the bias score in the open and secret conditions.

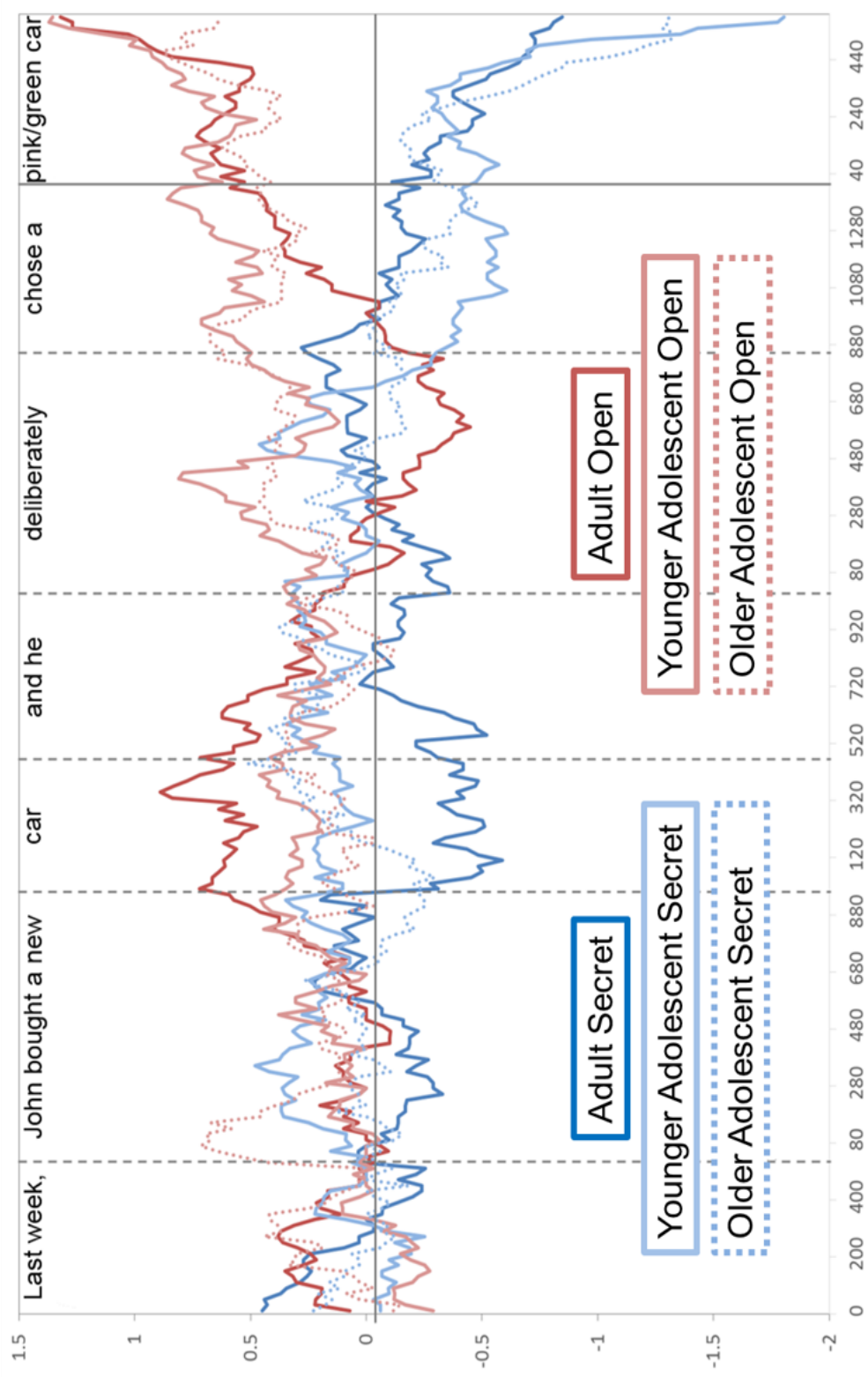
## 4.3 RESULTS

### 4.3.1 Eye-tracking results

**Table 13** shows the fixed and random effects for the model adopted for each time window, and **Figure 11** shows the average referent bias scores throughout the target sentence for each condition in the adult and young/older adolescent groups. Analyses during the ambiguous noun time window (e.g. "car") showed no significant effect of Age group (all  $ps > .22$ ) and only a marginal effect of Context ( $p = 0.07$ ). Crucially however, there was a significant interaction between Context x Age Group 1 (Adult vs. both adolescent groups). Follow up analysis of the effects of context in each group revealed that the Adult group employed significantly different viewing strategies in Open and Secret trials (Est. = .92, SE = .31,  $t = 2.97$ ,  $p = .005$ ), reflecting a bias to the open referent within an open context, and a bias to the secret referent within a secret context. In contrast, the younger and older adolescents as a group (Age group 1) did not show a significant difference in referential biases between the two context conditions (Est. = .07, SE = .25,  $t = 0.28$ ,  $p = .778$ ). The Context x Age Group 2 (young vs. older

adolescent groups) interaction was not significant, showing that neither adolescent group reliably predicted the intention-relevant referent in this time window.





**Figure 11** The average  $\log(\text{open}/\text{secret})$  score for each condition and age group. Note that the dashed and vertical lines indicate absolute onsets and average offsets of words in the target sentence.

**Table 13** Estimates and t-values for each time window of interest.

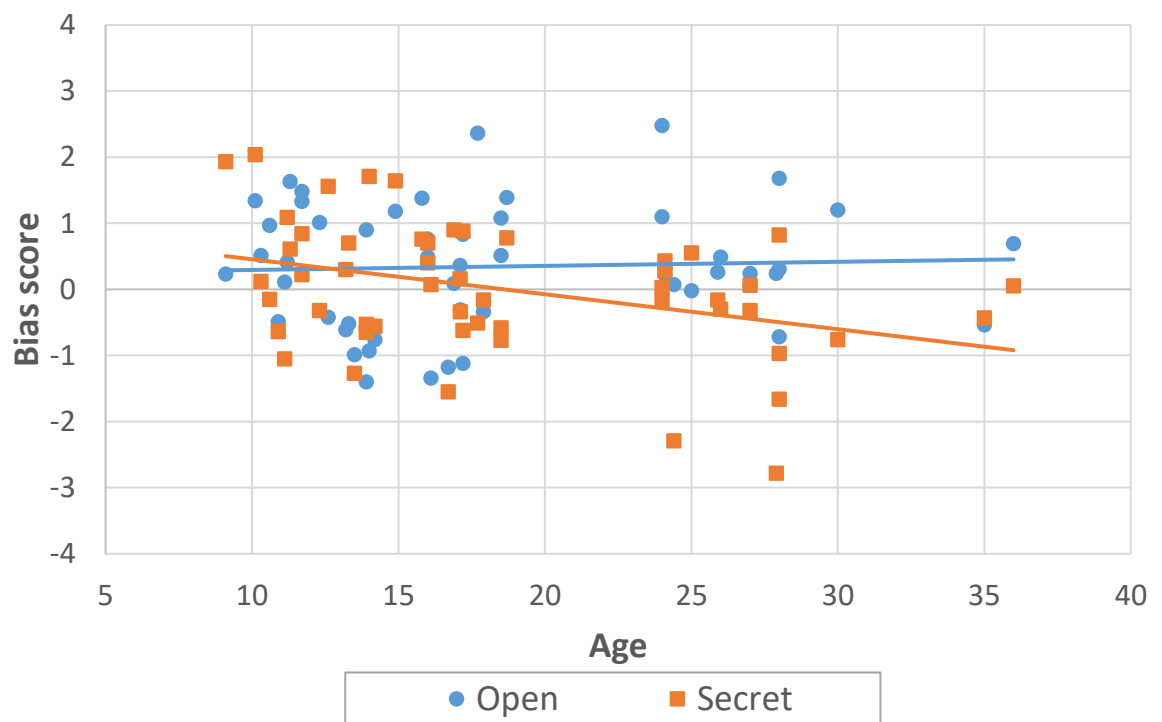
	Ambiguous noun		Post-ambiguous noun		Adverb		Transitive verb		Disambiguating noun	
	Estimate (SE)	t-Value	Estimate (SE)	t-Value	Estimate (SE)	t-Value	Estimate (SE)	t-Value	Estimate (SE)	t-Value
<i>Fixed effects</i>										
Intercept	0.17 (.14)	1.27	0.09 (.12)	0.80	0.13 (.17)	0.81	0.13 (.18)	0.74	0.04 (.16)	0.28
Context	0.35 (.19)	1.86	0.13 (.18)	0.72	0.34 (.20)	1.67	0.64 (.30)	2.12*	1.93 (.20)	9.59**
AgeGroup1	-0.16 (.14)	-1.13	-0.06 (.13)	-0.47	-0.30 (.16)	-1.91	0.00 (.18)	0.98	0.08 (.14)	*
AgeGroup2	0.11 (.25)	0.45	-0.08 (.22)	-0.36	0.01 (.27)	0.02	-0.18 (.31)	0.57	0.13 (.24)	0.53
Context:AgeGroup1	0.58 (.26)	2.23*	0.38 (.25)	1.53	-0.21 (.28)	-0.73	-0.51 (.30)	0.10	-0.23 (.29)	-0.81
Context:AgeGroup2	-0.14 (.46)	-0.30	0.40 (.44)	0.91	0.20 (.51)	0.39	0.62 (.54)	0.26	0.61 (0.51)	1.20
<i>Random effects</i>										
Participant (Var)	0.09		0.02		0.13		0.33		0.05	
Item (Var)	0.14		0.09		0.24		0.26		0.24	
Residual (Var)	6.80		6.20		8.10		7.43		7.05	

\* p < .05.

\*\* p < .01.

\*\*\* p < .001.

To further explore the age group difference found in this ambiguous region, post hoc Pearson correlation analyses were performed comparing age (as a continuous variable) with participants' bias score in the open condition and secret conditions separately.<sup>7</sup> This data is presented in **Figure 12**. In the open condition there was no correlation between age and the bias score [ $r(52) = .05, p = .748$ ]. However, age correlated with the bias score in the secret condition [ $r(52) = -.37, p = .007$ ]. This suggests that appropriate biases toward the secret referent in the secret condition became more likely and stronger with increasing age. Interestingly, age did not influence bias looks on the open referent in the open condition.



**Figure 12** Correlation between participants' age and bias score in the open condition and the secret condition.

Analyses in the time windows following the ambiguous noun ("and he") and adverb ("deliberately") showed no main effects or significant interactions (all  $ps > .1$ , with the exception of a marginal effect of Age group 1 in the adverb condition where  $p$

<sup>7</sup> Correlation analysis rather than regression (as performed Chapter 2) was appropriate here as we wanted to examine the relationship between two dependent variables that we measured and did not manipulate.

= .062).<sup>8</sup> In the transitive verb region (“chose”) analyses showed an overall significant effect of context but no effects of any Age group or any significant interactions (all  $ps > .1$ ). Here, participants showed an overall bias to the open referent within an open context, and a bias to the secret referent within a secret context, suggesting that by this point all participants had predicted the character’s intention-appropriate behaviour. This intention-appropriate bias continued into the disambiguating noun time window (“pink/green car”), showing a significant effect of context and no other effects or significant interactions (all  $ps > .24$ ).

To summarise, the eye-movement data suggest that adults made significantly different visual biases between the Open and Secret conditions from as early as the ambiguous term (e.g. “car”). They begin to correctly anticipate reference to the open referent in the open condition (i.e. the pink car) but crucially, they also correctly anticipate reference to the secret reference in the secret condition (i.e. the green car). This suggests that adults not only showed a sensitivity to the different contexts but used that information about the character’s intentions to make predictions. However, neither adolescent group seems to show sensitivity to the context until much later, during the transitive verb time window which precedes the disambiguating words, suggesting that they do not make inferences about the perspective of the characters early on, like adults do. Additionally, further analyses suggest that while the likelihood of making an intention-appropriate prediction of the character’s behaviour in the secret condition increases linearly with age, age does not correlate with the bias score in the open.

#### 4.3.2 Individual measures results

Results for one way ANOVA’s of all individual measures are shown in **Table 14**. Adults and both adolescent groups did not differ in any of the IC measures (all  $ps > .21$ ).

---

<sup>8</sup> The analysis here and as seen in Ferguson and Breheny (2011) suggests that adult participants do not have biases to either referent in each condition in the windows following the ambiguous noun until the transitive verb region. This is a pattern of bias that has also been in Ferguson & Breheny (2011). It seems that adults here are more sensitive to the task in that they consider the right answer early on and then look around to explore the other images. This is something similar to what we found in Chapter 2, where participants decide their answer early on and look at the items they will eliminate. This search behaviour may be encouraged by the use of the adverb, ‘deliberately’ which implies that the agent considered the alternative options.

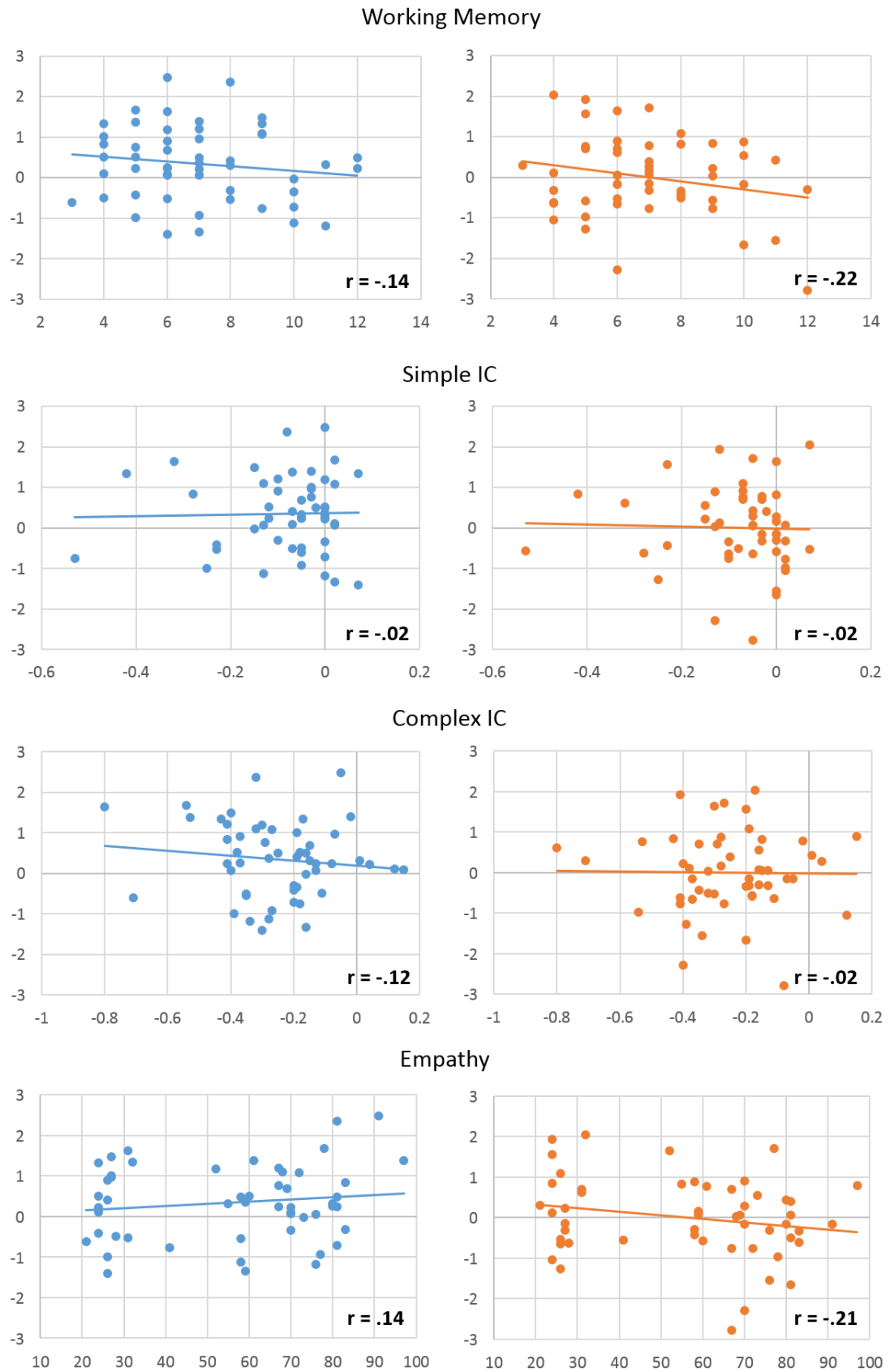
Analyses on the Digit Span score revealed a significant difference between age groups and Bonferroni corrected post-hoc comparisons showed that the younger adolescents performed worse than the older adolescent group ( $p = .05$ ) and adults ( $p = .003$ ), but that older adolescents and adults did not differ ( $p = .89$ ). For empathy, recall that one-way ANOVAs were performed using the EmQue-CA scores to examine age group differences between the younger and older adolescent groups and the IRI scores were used to compare the older adolescent group with adults. Analyses of the EmQue-CA scores showed a significant difference between age groups where younger adolescents had lower empathy scores than the older adolescent group ( $p = .028$ ). However, empathy (using the IRI scores) did not significantly differ between the older adolescent group and adults ( $p = .52$ ).

**Table 14** Means and standard deviations of all individual measure scores and their respective one way ANOVA results.

	Mean (SD)			ANOVA	
	Adult	Older Adolescent	Younger Adolescent	df	F
<i>Inhibitory control</i>					
<i>Accuracy</i>					
Simple Go NoGo	-0.05 (0.72)	-0.08 (0.14)	-0.11 (0.13)	2, 49	0.95
Complex Go NoGo	-0.21 (0.16)	-0.24 (0.15)	-0.32 (0.22)	2, 49	1.59
<i>RT</i>					
Simple Go NoGo	386.76 (19.43)	382.70 (47.25)	402.65 (55.68)	2, 49	1.01
Complex Go NoGo	429.15 (50.06)	425.31 (53.39)	453.38 (72.11)	2, 49	1.14
<i>Working Memory</i>					
Digit Span	8.00 (2.26)	7.28 (1.96)	5.59 (1.80)	2, 49	6.40**
<i>Empathy</i>					
EmQue-CA		29.13 (3.95)	26.35 (2.91)	1, 31	5.31*
IRI	71.88 (9.58)	69.28 (13.57)		1, 33	0.43

\*  $p < .05$ .  
 \*\*  $p < .01$ .  
 \*\*\*  $p < .001$

To further examine whether individual differences in IC , WM or empathy had an effect on anticipatory looks to the correct target, Pearson correlations were performed on participants IC (simple and complex), WM and empathy scores with the bias score in the open and secret conditions separately. Correlations were conducted for the ambiguous noun region (**Figure 13**), which is where the first evidence of context effects emerge in this study and Ferguson & Breheny (2011), and where age-group differences were found here. In addition, we conducted correlations for the entire anticipatory window (from the ambiguous word until onset of disambiguating word). Results showed that there were no significant correlations for any of the individual measures and the bias score in either condition in the ambiguous noun region (all  $p$ 's > .11) or the entire anticipatory region (all  $p$ 's > .37).



**Figure 13** Correlation between participants bias score (y-axis) in the open (blue) and secret (orange) condition in the ambiguous noun region and scores from each individual measure (x-axis).

To summarise, younger adolescents, older adolescents and adults did not differ in IC, however the younger adolescents exhibited lower WM and empathy than the older adolescents and adults. The older adolescents and adults did not differ in any of the individual measures. Moreover, none of the individual measures significantly associated with visual biases in the eye-tracking data.

#### **4.4 DISCUSSION**

The current experiment had two main objectives, i) to replicate the results of Ferguson & Breheny (2011) in adult participants, and ii) to examine whether younger and older adolescents can use information about a character's basic preferences and higher-order desires to make complex ToM inferences and predict that character's subsequent behaviours as rapidly as adults. Participants' eye-gaze was tracked while they listened to short stories that first introduced a property of a character then described that character performing an action that was consistent with their intentions. In the open condition, the character's basic preferences and high-order desires matched, whereas in the secret condition the character's basic preferences and high-order desires were in conflict. A visual display was presented during the second sentence that included images that matched the character's action. To explore effects from individual differences participants also undertook standard tests of IC, WM and an Empathy Quotient test.

Results from the secrets task showed that adults anticipated the appropriate target long before disambiguating information was presented (i.e. in the ambiguous 'car' region) in both the *open condition*, when there was no conflict between the character's basic preferences and high-order desires, as well as the *secret condition*, when there was a conflict between the character's basic preference and high-order intentions. This pattern is in line with Ferguson & Breheny (2011)'s findings. Crucially, younger and older adolescents showed a delay in anticipating the target in both conditions, as distinct gaze biases for open and secret conditions only emerged in the region immediately preceding the disambiguating information (i.e. during 'chose'). Moreover, further analyses during the ambiguous noun region suggest that the likelihood of appropriate biases to the secret referent in the secret condition increases linearly with age, but age does not correlate with the bias score in the open condition.



Additional analyses of individual difference characteristics showed no age-related differences in participants' inhibitory control performance. However, younger adolescents performed significantly worse than older adolescents and adults on tests of WM and empathy; older adolescents and adults did not differ. Thus, adolescents were delayed relative to adults at anticipating the character's actions based on intentions despite comparable IC, and younger and older adolescents performed comparably despite age-related differences in WM and empathy. None of the individual measures correlated with any of the eye-tracking data.

The results here are similar to those in Chapter 2 (and Symeonidou et al., 2016) where we found age-related differences between adolescents and adults. However, one major difference in the current study is that individual measures, were not associated with performance in the secrets task. Notably, in contrast with the DT task where IC accounted for some of the variance, IC does not seem to be factor in this task as a between-groups difference in secrets performance was not reflected in either of the IC tasks. In other words, adolescents have an adult-like IC performance but yet still have a difficulty in the secrets task. This suggests that perhaps age-related differences are not attributed to poorer IC, WM or empathy in this particular task, rather other factors seem to be affecting their online use of ToM.

These findings are not in contrast to previous findings that suggest that these factors give rise to individual differences in ToM abilities. Rather, these findings are consistent with our assumption and also with Ferguson & Breheny (2011) that EF demands in the secrets task are lower than the DT. Note that the suggestion is not that IC is not involved in the secrets task at all. Rather IC is involved in the secrets condition but I suggest it is a different type of IC than what is involved in the DT. Namely, the information one has to inhibit during the secrets condition is not privileged information that is salient just to you. Instead, the information here is shared knowledge. During the DT task and the FB task one has to inhibit one's own privilege knowledge to be successful in the task. Some evidence to support this idea comes from the fact that IC correlated in the DT (Chapter 2) and FB task (Chapter 3), but it does correlate with the secrets task. In terms of WM, it could be argued that representing information about higher order mental states requires more WM. However, we know of no research that establishes

this and the current results do not correlate with individual measures of WM and we leave it for future research as to why processing higher order mental states might require more WM. Therefore, although IC, WM or empathy might not be the main constraining factors, further research is needed to identify whether IC, WM or empathy account for any variance in such a task by including a behavioural measure which was not present here.

In this experiment, we eliminated the possible factors that are argued to interfere with perspective-taking in the DT (i.e. conflict from linguistic input or knowledge from the self-perspective). Nevertheless, in comparison to adults, all adolescents showed a delay in generating expectations based on mental state information. In Chapter 3 we found that adolescents are sensitive to others' perspective as early as adults when: i) relevant information about the character's beliefs are salient and more the focus of attention in the FB trials, ii) participants do not have to plan their own action based on someone else's belief iii) there is no ambiguity to resolve, but the participant still maintains a more informed perspective. The combination of these factors, I argue, make it easier to integrate mental state information on line to predict the character's actions. In the secrets task all of these possible constraints were avoided as well as an additional possible constraint, i.e. conflicting knowledge from the self-perspective. Note, as discussed above the assumption that EF demands are lower in the current task is consistent with the individual measures results where they do not predict performance during the secrets task. This assumption is also supported by the fact that adult participants perform successfully in the secrets task whereas in the DT they do not. Despite this, adolescent participants still suffered a delay in comparison to adults. What then is a factor constraining adolescents ToM processes during the secrets task?

The task requires participants to make a prediction about the character's action using information about the character's mental state. In the open condition, the character's basic and high-order desires are in sync which means that either inference (the basic preference or intention) a participant makes will be correct. During a secret trial however, the character's basic belief is in contrast with their high-order intentions and participants have to consider both mental states that are in conflict. More specifically, the participant has to consider that the character can do one of two possible

actions. Choosing to buy the pink car (see example in **Table 11**) is in accordance with the character's basic preferences but seems socially unacceptable given the character's higher-order preferences. On the other hand, choosing to buy the green car, while it is not the character's basic preference, is more socially acceptable according to the character's desires. The participant needs to consider if the character is open or embarrassed/self-conscious about their preferences. Perhaps then although IC demands are less demanding (and perhaps different as suggested above) here, successful performance on this task requires rapid integration of more complex ToM inferences, which might be what adolescents struggle with during ToM processing.

Recall that while age correlated with the bias score in the secret condition (i.e. appropriate biases toward the secret referent became more likely and stronger with increasing age), age did not have an effect in the open condition. This could suggest two possible explanations. One possibility is that younger participants have difficulty with the secrets condition in particular because it requires inferences about situations in which a character has to resolve competing desires. It might be that they are not used to thinking about how to resolve a conflict between what you want to do and how you want other people to perceive you. Previous research on ToM development in 8-13 year olds has shown that self-reported social experiences (specifically loneliness and peer rejection) accounted for some of the variance in poorer performance during a ToM task (Devine & Hughes, 2013). While this report only focused on loneliness, social adequacy and estimates of peer status (peer rejection) it suggests that social experiences are a factor in ToM abilities. Including a questionnaire on social experiences for future research could shed light on how much social experiences play a role in ToM development.

A second possibility could be that when the level of ToM inferential complexity is higher, adolescents' performance in the open condition is facilitated by a low-level bias to the open target from the facts stated in the context sentence; i.e. a participant hears 'pink' in the context and the target is 'pink car' in the open condition whereas it is 'green' in the secret condition. This is discussed in Ferguson & Breheny (2011), and was found to be a factor in their follow up experiment. Therefore, it is possible that the low-level bias helped adolescents in the open condition, but conflicted in the secrets

condition, hence explaining why age correlated with the secret condition only. Note that this interpretation is consistent with adolescents not being as competent at higher-order ToM inferences across the board, and instead relying on low-level cues. One way to address this question is to run a follow-up study as in Ferguson & Breheny (2011) to examine whether this low-level bias actually aids adolescents' performance or if adolescents truly find the open condition as difficult. The follow up involves the same stimuli and design with the exception of the valence of the intentionality adverb in the target sentence which makes participants set up expectations about the character's subsequent behaviour that are in contrast to the preceding context. For example, in the open condition if the preceding context involved *Tom* always telling people that his favourite colour is pink, the target sentence would be that he *surprisingly* chose the green car. Similarly, in the secret condition, while the context would indicate that Tom does not want people to know that he likes the colour pink, in the target sentence Tom *surprisingly* chooses the pink car. If adolescents are struggling with ToM and, therefore, fall back to these the low-level bias created in the current experiments to help their performance in the open condition then I would expect that in this follow up study adolescents would perform just as poor in both conditions. An alternative is that these low-level biases are just stronger in adolescents than mental state biases in which case you would see a reversed effect of pure priming.

One additional factor to consider here to explain the poorer performance in adolescents overall, might be the emotional component related to the characters' conflict between basic and higher-order preferences. For example, while a character loves junk food, she is embarrassed for people to know that. Understanding the desire for this character in order to predict their behaviour requires some understanding of emotions; the ability to infer what the affective state of that person is rather than just their belief or motivation. Moreover, one needs to understand how that emotion can impact his/her motivation or desire. Recall in Chapter 1 I presented a recent model which proposes that ToM is comprised of two main components: i) cognitive ToM which refers to the ability of inferring others beliefs, and ii) affective ToM which refers to the ability of inferring others affective state (Shamay-Tsoory, Harari, Aharon-Peretz, & Levkovitz, 2010). As I reviewed in Chapter 1, studies have suggested that adolescents'

perform poorly on affective ToM tasks in comparison to cognitive ToM tasks (Sebastian et al., 2012; Vetter, Altgassen, et al., 2013). Sebastian et al., (2012) observed that adolescents (11-16) made significantly more errors than adults (24-40) in cartoon scenarios that involved affective ToM where participants had to infer how one story character would react to the other character's affective state. Conversely, there were no age-group differences in cartoons that involved cognitive ToM where participants had to infer how one character would react to the other based on beliefs and intentions. Previous research has also found that children fail to correctly judge emotional reactions of someone who has a false-belief (Hadwin & Perner, 1991). It could be then that ToM abilities in adolescents suffer when emotion is involved either because it is harder for them to recognize emotional states or integrating how the affective state will impact someone's desire. Note that this hypothesis is only an additional suggestion on the main explanation of the data about the complexity of ToM given above. Moreover, this suggestion is a speculative one as we could not directly compare all of the age groups with the same empathy scores. Further investigation is required to disentangle how much emotion plays a factor in ToM use during adolescents and I put forth this proposal here as I believe it is a fruitful direction for further research when considering all of the adolescent research as a whole.

In sum, these results provide further evidence for age-related differences between adolescents and adults in their online use of ToM. Specifically, when reasoning about the character's basic preference and high-order intentions, our results suggest that adolescents may not be able to use information about others' mental states for language comprehension as quickly as adults. Although WM and IC have been shown to be factors in the application of ToM in previous studies, the current results indicate that these are not the only factors that play a role in the online application of ToM. Combining these results with the previous findings, it seems that while adolescents can consider other people's perspective as early and as spontaneously as adults when they have no explicit reason to do so, when ToM demands are higher and require integration of more complex inferences online, adolescents seem to struggle, even when EF demands are arguable low. What might be more 'complex' here for adolescents could simply be the fact that inferences are based on higher-order mental states. As

suggested, it might be additionally more complex because there are emotional components they have to consider and integrate. Although further investigation is required to explore this hypothesis, I propose that this is another possible ability that is still maturing during adolescence.

## 5 CHAPTER 5: IRONY PROCESSING AND ADOLESCENTS

---

### 5.1 INTRODUCTION

#### 5.1.1 Theory of mind and adolescents

Results in Chapter 2 revealed age-related differences between adults and adolescents above and beyond inhibitory control demands that I attributed to adolescents' lesser ability detect and integrate relevant mental state information rapidly from non-salient cues. I tested this hypothesis in Chapter 3 using a false-belief paradigm that involved the same type of inferentially complexity as the DT in Chapter 2 but provided more salient perspective cues, eliminated linguistic conflicts, and reduced EF complexity. Results suggested that adolescents can spontaneously infer other people's perspective during false-belief reasoning when available cues are salient, and EF demands are lower. However, in Chapter 4 results suggested that when the task involves a greater degree of ToM inferential complexity, adolescent participants perform less well in comparison to adults, even when EF demands are low. Specifically, when reasoning about the character's basic preference and high-order intentions, results suggest that adolescents may not be able to use information about others' mental states for language comprehension as rapidly as adults. Combining these results together, it seems that when ToM demands are higher and complex integration is required online, adolescents do not perform at an adult level. In the current chapter, I further explore adolescents' advanced ToM inferential abilities by investigating irony comprehension. Specifically, in the current study, I explore the time-course of irony comprehension through event-related brain potentials (ERPs) for the first time in adolescents and investigate any age-related differences between adolescents and adults.

Irony comprehension requires complex inferences about a speaker's mind (Pexman, 2008). Correct interpretation during communicative exchanges often requires mental state attribution as the expressed meaning is not always the intended one (Filippova, Astington, Filippova, & Astington, 2008; Happé, 1993). For example, consider a context where a singer tells her colleague: "That was a superb performance tonight", after a performance in which both singers forgot their words, and are aware it was a

poor performance. The singer is clearly being ironic. Successful interpretation of the irony requires the listener to go beyond the literal meaning that is conveyed and understand the speaker's intentions. To fully understand irony, the listener needs to recognize that the speaker did not intend for the listener to take meaning of the utterance literally and that the speaker's belief differs from the literal meaning of the utterance; i.e. in the statement above the speaker's true belief is that the performance was terrible (Nicholson, Whalen, & Pexman, 2013; Pexman & Glenwright, 2007). Moreover, the listener has to understand that the speaker intentionally wants the listener to recognize this falsehood and crucially, the listener needs to understand the speaker's attitude and appreciate why the speaker chose to communicate in this manner (e.g. to tease, mock, joke) (Filippova et al., 2008; Nicholson et al., 2013; Pexman & Glenwright, 2007). Further, it is argued that being able to understand the speaker's attitude and intention for the listener's beliefs about the statement is what distinguishes irony from deception (Demorest, Meyer, Phelps, Gardner, & Winner, 1984; Demorest, Silberstein, Gardner, & Winner, 1983; Sullivan, Winner, & Hopfield, 1995; Winner et al., 1987; Winner & Leekam, 1991).

### **5.1.2 Theory of mind and irony processing**

Previous neuroimaging studies have provided evidence which suggests that irony comprehension involves the ToM network at the neural level (Shibata, Toyomura, Itoh, & Abe, 2010; Spotorno, Koun, Prado, Van Der Henst, & Noveck, 2012; Wang et al., 2006). Shibata and colleagues (2010) examined the neural substrates involved in irony comprehension by presenting short stories to adult participants with either an ironic, literal or unrelated final sentence. Participants then had to indicate whether they thought the sentence was ironic or not by pressing a button. Compared to the non-ironic sentences, the ironic sentences were associated with higher activation in brain areas that are thought to be part of the ToM network (the right medial prefrontal cortex, the right precentral and the left superior temporal sulcus), suggesting a relationship between ToM and irony comprehension. Similarly, an fMRI study by Spotorno, Cheylus, Van Der Henst & Noveck (2012) investigated whether the ToM network is involved during irony processing in adult participants. The study found that ironic statements



consistently recruited the ToM network more so than literal statements. They showed that the ToM network is activated when a participant is processing verbal irony.

In line with neuroimaging data, behavioural research has shown that ToM processing plays a significant role in irony comprehension. Filippova and Astington (2008) measured children's (5, 7, and 9 year olds) ability to attribute others' mental states through irony comprehension. Participants were first presented with a scenario that involved a character being the subject of an ironic or literal comment, which was followed by comprehension questions in order to establish whether participants paid attention to the story. The researchers then asked participants questions to examine if they recognized i) the discrepancy between the literal meaning and what the character actually meant, ii) the speaker's actual belief and what their communicative intentions were (i.e. does the speaker want the listener to believe that the speaker thinks that) and iii) the speaker's motivation and attitude. Moreover, participants' advanced ToM abilities (second-order beliefs, intentions, feelings and motivations) were measured with tasks such as the "strange stories" task (Happé, 1994) and the "faux pas stories task" (Banerjee, 2000; Baron-Cohen, O'Riordan, Stone, Jones, & Plaisted, 1999) in order to assess whether ToM abilities are a contributing factor to successful irony comprehension. Results from this study showed a positive correlation between participants' performance in the advanced ToM tasks and their ability to successfully identify the speaker's communicative intention in the irony task, suggesting that ToM is involved during children's irony comprehension.

More recently, Spotorno and Noveck (2014) found an association between adult participants' irony comprehension and their social abilities, as measured by the Social Skill subscale in the Autism Spectrum Quotient (AQ questionnaire; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). Specifically, participants who had lower AQ scores and were more socially inclined were more likely to use a story's negative event as a cue to process an ironic utterance as quickly as its literal counterpart.

### 5.1.3 Irony processing in development

Numerous studies have investigated how the processing of irony develops. On the one hand, a body of research has found that children as young as 6 begin to recognize irony and that speaker does not believe the literal meaning of their utterance (Ackerman, 1981, 1983; Andrews, Rosenblatt, Malkus, Gardner, & Winner, 1986; de Groot, Kaplan, Rosenblatt, Dews, & Winner, 1995; Dews et al., 1996; Filippova et al., 2008; Hancock et al., 2000; Harris & Pexman, 2003; Nakassis & Snedeker, 2002; Pexman & Glenwright, 2007; Winner & Leekam, 1991). However, there is evidence to suggest that full appreciation of irony continues to develop during middle childhood (Filippova et al., 2008; Pexman et al., 2005) and into adolescence (Demorest et al., 1984).

Ackerman (1983) tested 6-8 year olds' understanding of ironic utterances and proposed that the development of irony comprehension is composed of two stages (Ackerman, 1983; Hancock et al., 2000). Children are first able to understand that the speaker's belief is not the literal interpretation of the statement and to reject the literal meaning of an utterance and at a second stage, children are able to infer the speaker's intention and social purpose.

Demorest and colleagues (1984) investigated children's (aged 6 and 9), young adolescents' (aged 13) and adults' ability to comprehend sincere, deceptive and sarcastic utterances. In line with Ackerman's (1983) view, Demorest and colleagues (1984) found the following age-related differences. 6-year olds interpreted all three types of utterances as sincere. The 9-year olds and 13 year olds group tended to interpret all utterances as deceptive, in comparison to adults who distinguished between deceptive and ironic remarks reliably. The authors suggest that this is because while children/young adolescents appreciate that the speaker's utterance intentionally differs from her belief, they do not reliably infer speaker's intention and purpose which is what distinguishes irony from deception. This suggests that even 13 year olds do not distinguish irony from deception at adult levels (Demorest et al., 1984, 1983; Filippova et al., 2008).

Similarly, in Filippova and Astington (2008) (described above), results showed that children who rejected the literal meaning of the sentence did not understand the speaker's communicative intent. The authors argue that the understanding of the full

communicative intention emerges later and only then do children start to recognize speaker attitude.

Age-related functional differences in irony comprehension during adolescence have also been found. In an fMRI study, Wang and colleagues (2006) studied whether there are developmental changes in pragmatics at the neural level. Specifically, they investigated whether the neural circuit involved in interpreting communicative intent differs in children/adolescents (9-14 years old) and adults (23-33 years old). Their results showed that children and adolescents activated the mPFC, left IFG and the right pSTS more than adults during the ironic condition versus the non-ironic condition, indicating a different neural process between adolescents and adults. These results further support the hypothesis of a possible developmental trend in behaviour beyond middle childhood and into adolescence during irony comprehension. Taken together, these findings indicate that full appreciation of irony continues to develop in middle childhood and warrant further investigation into the adolescent years.

#### **5.1.4 Irony processing and P600**

Research using fMRI has provided evidence about the brain areas that might be involved during irony processing, however due to its low temporal resolution fMRI does not provide any insight into how irony comprehension unfolds in real time. As such, while there is evidence that brain activation differs between adolescents and adults during irony comprehension, no studies have investigated whether these differences are reflected in how irony is processed in real time.

Electroencephalogram (EEG) measures the electrical activity in the brain. However, in its raw form it is difficult to detect and measure specific neurocognitive processes (Luck, 2014). In order to extract specific responses that are related to certain neural processes, the EEG needs to be time-locked to the presentation of the stimulus. ERPs are time-sensitive measurements that provide continuous information on cognitive processes underlying a response. Research on irony comprehension using ERPs has reliably found that adult participants showed an enhanced late positivity (also known as the P600), that has an onset around 500-600ms after stimulus onset, for ironic

sentences compared to literal sentences (Regel, Coulson, & Gunter, 2010; Regel, Gunter, & Friederici, 2011; Regel, Meyer, & Gunter, 2014; Spotorno, Cheylus, Van Der Henst, & Noveck, 2013). However, it is important to note that the functional significance of the P600 component is still debated as it has been associated with a number of linguistic phenomena such as semantic reversal anomalies (e.g. "That cat that fled from the mice", van Herten, Kolk, & Chwilla, 2005), semantic illusory anomalies in a discourse context (e.g., "How many animals of each sort did Moses put on the Ark"?, Nieuwland & Van Berkum, 2005), humour comprehension (Coulson & Kutas, 2001; Shibata et al., 2017), indirect requests (Coulson & Lovett, 2010), and metaphors (Coulson & Van Petten, 2002; De Grauwe, Swain, Holcomb, Ditman, & Kuperberg, 2010).

### 5.1.5 Empathy and Irony processing

From the developmental research on irony comprehension there seems to be a consensus that what develops last is the ability to understand the speaker's attitude and motivation behind the ironic remark. It has been suggested that fully understanding irony, including understanding of speaker's intent and attitude, requires some understanding of emotion and the ability to infer how a person is feeling. As I presented in Chapter 1 and 4 Shamay-Tsoory, and colleagues' (2010) model proposes that ToM is comprised of cognitive ToM and affective ToM (Shamay-Tsoory et al., 2010). Affective ToM is suggested to be the integration of cognitive ToM and empathy (Sebastian et al., 2012; Shamay-Tsoory et al., 2010).

Shamay-Tsoory and colleagues (2005) found that patients with lesions in the right ventromedial prefrontal cortex (vmPFC), which is suggested to be required for affective but not cognitive ToM, were able to perform well on false-belief tasks but not on irony and "faux-pas" tasks. The authors explained these results by suggesting that understanding irony requires the ability to infer the emotional states of others whereas false-belief tasks do not. These findings then suggest that, empathy also plays a role in irony comprehension.

In line with this hypothesis, Nicholson, Whalen & Pexman (2013), found that empathy development correlated with irony comprehension in 8 and 9-year-old

children. Notably, as reviewed in Chapter 1 and 3, an ERP study (Ferguson, Cane, et al., 2015) has shown that individual differences in empathy are associated with participants' abilities to integrate others' beliefs online. Ferguson and colleagues (2015) found that participants with high levels of empathy are better at making inferences according to someone else's (false) belief FB (which modulated the amplitude of the N400 effect). Given these results, I want to investigate whether the ability to understand irony is modulated by individual differences in empathy skills in adolescents and adults.

#### 5.1.6 Current study

Evidence shows that the ToM network develops through adolescence and that same network is involved in irony comprehension. In addition, fMRI evidence indicates that adolescents activate this network differently from adults during an irony task. Behavioural data indicates that while children between 6-10 years of age begin to show some appreciation of the communicative intent in an ironic remark (Filippova et al., 2008; Pexman & Glenwright, 2007), full appreciation of intent and speaker attitude might be developing beyond middle childhood and into adolescence. In this study, I aim to explore how the adolescents' brain processes irony in real time and compare that to adults. Moreover, I want to investigate whether individual differences in empathy are associated with irony processing.

In the current experiment, I adopted Spotorno and colleagues (2013) paradigm to examine the time-course of irony comprehension in adolescents, in comparison to adults. I recorded participants' electrical brain responses as they read seven-sentence stories in which the target sentence was either literal or ironic (see **Table 15**). I chose to use this paradigm for two main reasons. Firstly, this paradigm has been used in an fMRI study (Spotorno et al., 2012) where results showed that the ToM network is recruited while a participant is processing the ironic remark versus a literal one. As such, we can be confident that, at least in adults, the stimuli include a rich enough context that engage the ToM network. This is crucial, since the current study aims to explore the development of ToM as revealed in irony comprehension. Secondly, this same paradigm has been used in an ERP study with adults (Spotorno et al., 2013). Thus, we have clear

predictions for the time-course of irony comprehension in adults and we can compare when and if processes diverge in adolescents.

To examine whether empathy is associated with irony comprehension, participants' empathy was measured through the Interpersonal Reactivity index (IRI; Davis, 1983). Moreover, to ensure that all participants had a basic understanding of irony comprehension, participants also completed a simple irony comprehension questionnaire based on the irony comprehension task of Wang and colleagues (2006).

To sum up, this paper aims to i) replicate the adult results from Spotorno and colleagues (2013), ii) examine how adolescents' brain processes irony comprehension in real time in comparison to adults and iii) investigate whether individual differences in empathy affect irony processing. Based on previous findings (e.g. Spotorno et al., 2013), we expect adult participants to show an enhanced P600 in the ironic condition in comparison to the literal condition. If the age-related differences previously found in ToM processing (Chapter 2 and Chapter 4) are due to integrating information about other's mental states online, which as discussed above is involved in irony comprehension, then we expect to see differences in adolescent participants. Finally, if affective ToM is engaged during irony comprehension as previously suggested, then we expect that participants with higher empathy score will show a larger P600 effect in irony comprehension.

## **5.2 METHOD**

### **5.2.1 Participants**

A total of fifty participants took part in this study: twenty-four adults (24-34 years old,  $M = 28.21$ ,  $SD = 2.67$ ) and twenty-six adolescents (11-18 years old,  $M = 14.62$ ,  $SD = 2.28$ )<sup>9</sup>. All participants were native English speakers. Participants were recruited as in Chapters 2-4. Data from three adolescents and one adult were excluded from the analysis due to excessive artifacts (>50% epochs rejected in one or more conditions). All adult participants and the parents/guardians of all adolescent participants were given

---

<sup>9</sup> Note that younger adolescents were not used in the current task as in Chapter 2 and 4 due to the difficulty of the task; it was very long, participants are required to read long stories, and are not allowed blink for certain parts of the task.

information sheets prior to the study and informed consent was obtained. This study was approved by the UCL Research Ethics Committee.

## 5.2.2 Stimuli and design

### 5.2.2.1 ERP Experiment

The task design and stimuli were based on a previous study by Spotorno and colleagues (2013). The experiment had a 2 x 2 mixed design, with *Story type* (Ironic vs Literal) as the within-participant factor and *Age group* (Adult vs Adolescent) as the between-participants factor. All sixty sets of stories in the experimental conditions from Spotorno and colleagues (2013) were translated from French to English (see **Appendix IV** for all of the experimental items). The stories were translated as closely as possible to the original ones but minor changes were made where necessary to ensure that they sounded natural in English and were more culturally appropriate. For example, stimuli that referenced French cities were changed to English ones.

**Table 15** shows an example of a story for each condition along with both types of filler stories. Every story was seven lines long and each story described an everyday situation which involved two characters who knew each other. The first three sentences of the story introduced the situation and the characters. The fourth and fifth sentences explained how the story developed in either a positive (literal condition) or negative (ironic condition) way. These are the only two sentences that differed between each set of stories. In many cases only sentence four differed. The sixth sentence was the target sentence which was identical across the ironic and literal conditions and the critical word (e.g. *superb*, see **Table 15**) allowed the participant to determine whether the sentence was ironic or not. What made the target sentence ironic or literal was the context difference in sentences four and five. It is worth noting that in the Spotorno and colleagues' (2013) study the critical word was always at the end of the sentence. However, in order to avoid potential wrap-up effects, we extended the target sentences in the present study to ensure that the critical word was not sentence-final. Finally, the last and seventh sentence concluded the story in a way that made sense for both the ironic and literal conditions.

**Table 15** Examples stories for each condition and each type of fillers. Critical word is underlined.

Condition	Example
Ironic	<p>Janet and Leanne sing in the same opera.</p> <p>The evening of the first performance, they meet at the theatre.</p> <p>The show starts right on time.</p> <p>During the performance, they forget their words and sing out of key.</p> <p>After the show, Janet tells Leanne:</p> <p>"Our performance was <u>superb</u> tonight."</p> <p>While they are removing their make-up, the two girls continue to talk about the show.</p> <p>Question: Are Janet and Leanne theatrical technicians?</p>
Literal	<p>Janet and Leanne sing in the same opera.</p> <p>The evening of the first performance, they meet at the theatre.</p> <p>The show starts right on time.</p> <p>The performance is excellent and the singers are applauded for a long time.</p> <p>After the show, Janet tells Leanne:</p> <p>"Our performance was <u>superb</u> tonight."</p> <p>While they are removing their make-up, the two girls continue to talk about the show.</p> <p>Question: Are Janet and Leanne theatrical technicians?</p>
Fillers (Decoy)	<p>Mathew is moving out and has to move a heavy and very fragile mirror.</p> <p>He asks Paul to help him.</p> <p>Paul has nothing to do and comes to help him straight away.</p> <p>As soon as they lift the mirror, it breaks into a thousand pieces.</p> <p>Mathew tells Paul:</p> <p>"This was the only valuable thing in my place."</p> <p>A few days later, Mathew has a housewarming party at his new place.</p> <p>Question: Did Mathew and Paul move the mirror without trouble?</p>
Fillers (Positive)	<p>Jeremy promised his son to build him a tree house.</p> <p>He bought chestnut wood to make it.</p> <p>He worked an entire afternoon to build it.</p> <p>Once finished, the tree house is very solid and well-built.</p> <p>His son is very happy and tells his father:</p> <p>"Come play with me in the tree house."</p> <p>Both of them played together the entire weekend in the new tree house.</p> <p>Question: Was the tree house well-built?</p>



As the ironic story type always involved a negative context (e.g. a terrible performance), *decoy* filler stories were included with a negative context that did not lead to an ironic remark (see **Table 15** for example). This prevented the participants from using the negative contexts as a cue to an ironic condition. There were 30 decoy stories that were created with the same structure as the ironic and literal stories where a negative event occurs in sentences four and five. To counter-balance for these, an additional 30 fillers that had a positive context with a positive remark in sentence 6 were also included.

In sum, each participant was presented with 30 ironic stories, 30 literal stories, 30 decoys and 30 positive fillers. One version of each experimental story was assigned to one of two lists and participants were randomly assigned to one of two lists. The 30 decoys and 30 positive fillers were the same in each list but were interspersed randomly among the 60 experimental stories. A Yes/No comprehension question was presented after one third of the items to ensure participants were paying attention to the stories. The questions never referred to the target sentence. No feedback was given to the participants for their responses to the questions.

### 5.2.3 Questionnaires: empathy and irony comprehension

In order to check that all participants had a basic understanding of irony, a simple irony comprehension questionnaire was used that was based on the irony comprehension task of Wang and colleagues (2006) (see **Table 16**). Participants were presented with three-sentence stories. Each story differed in the second sentence where the context made the last sentence ironic or literal. Participants were then asked to respond with 'yes' if they thought the comment the speaker made was sincere and should be taken literally or 'no' if they thought the comment was sarcastic and the speaker meant the opposite of he or she said. There were 20 sets of short stories. One version of each story was assigned to one of two lists and participants were randomly assigned to one of two lists.

**Table 16** Example stories for irony comprehension questionnaire.

	1. Ben and Diana are blowing up balloons
Ironic	2 a) Ben keeps blowing his until it pops.
Literal	2 b) Ben keeps blowing his until it is huge.
	3. Diana says, “Nice going!”

The Interpersonal Reactivity index (IRI; Davis, 1983) was used as a measure for empathy (See Chapters 3 and 4 for the details of the questionnaire).

#### 5.2.4 Procedure

Participants were tested individually and completed the ERP experiment and the questionnaires (the IRI and irony questionnaire) in one session of approximately one hour and a half. The irony comprehension questionnaire was always completed after the ERP experiment so that participants were not primed to the experimental manipulation in the ERP experiment. Stimuli for the ERP experiment were presented using Presentation 18.3 (Neurobehavioral Systems, [www.neurobs.com](http://www.neurobs.com)) and the IRI and irony comprehension questionnaires were presented through Qualtrics.

The electroencephalographic (EEG) procedure was explained to each participant before the main experimental task. Once they were seated in a booth and the electrodes were applied they read the instructions from a computer screen. There were four practice trials before the main experiment to familiarise participants with the task. The main irony task was divided into four blocks of 30 stories each.

Each trial began with a white fixation presented at the centre of a black screen to indicate the start of a new trial. Afterwards, each sentence of the story was presented one by one on the screen. Participants read them in a self-paced manner such that the sentence remained on the screen until they pressed a key to move to the next sentence. There was a 500ms interval between the presentation of each sentence. In the literal and ironic conditions, the target sentence was always presented word-by-word. However, in the decoys and filler stories one in seven sentences was chosen randomly to be presented word-by-word in order prevent participants from using the word-by-word presentation as a cue about the aim of the experiment. For sentences that were

presented word-by-word, the stimulus onset asynchrony (SOA) was 650ms. At the end of one third of the trials, a comprehension question appeared on the screen and participants responded with one of two buttons on the keyboard. Participants were instructed to read at a normal pace, to pay attention to the stories and respond to the questions as accurately as possible. The presentation order of the stories was pseudorandomized so that 15 ironic, 15 literal, 15 decoys and 15 filler stories were presented in each half of the experiment.

### 5.2.5 Electroencephalogram (EEG) Recording

EEG data was recorded using 32 electrodes fixed to an elastic cap (Biosemi; <http://www.biosemi.com/headcap.htm>) using the international 10-20 system. All electrode offsets were kept within  $\pm 20$  mV and EEG recordings were sampled at 2048Hz. Offline, all electrodes were re-referenced to the average of left and right mastoids and down-sampled to 512Hz. An external electrode below the left eye was used to measure blinks and vertical eye movements (VEOG) and another external electrode on the right canthus was used to measure horizontal eye movements (HEOG). EEG activity was band-pass filtered (0.01-30Hz).

### 5.2.6 Data analysis

#### 5.2.6.1 ERP analysis

The continuous EEG data was segmented into 1250ms epochs (from 250ms prior to the onset of the target word to 1000ms post-onset) and a baseline correction was computed 250ms prior to the onset of the target word. Artefact rejection was conducted for all trials. Trials contaminated by blinks, drifts and other artifacts were discarded and not included in the analyses. This resulted in an average trial-loss of 13.3%.

For the analysis, 12 electrodes were chosen to define six scalp regions of interest which were distributed as follows: left-anterior (F3, C3); midline-anterior (Fz, Cz); right-anterior (F4, C4); left-posterior (P3, O1); midline-posterior (Pz, Oz); right-posterior (P4, O2). Following Spotorno and colleagues (2013), statistical analysis was conducted on

mean ERP amplitude in the time interval from 500ms to 800ms in order to analyse experimental effects on the P600.<sup>10</sup> Mixed repeated measures ANOVA were performed with *Story type* (Ironic, Literal), *Laterality* (Left, Midline, Right) and *Anteriority* (Anterior-central, Parietal-central) as within-subject factors and *Age group* (Adolescent, Adult) as a between-subject factor was performed. Note that in Spotorno and colleagues (2013) only nine electrodes were used for the analysis (frontal: F3, Fz, F4; central: C3, Cz, C4; parietal: P3, Pz, P4). To ensure that the additional electrodes did not have an effect on the results, we conducted the same analysis with the mentioned 9 electrodes. As the analysis revealed similar results they are only reported in **Appendix IV**.

#### 5.2.6.2 Questionnaires analysis: empathy and irony comprehension

Pearson's correlations were computed to examine the effect of individual differences in empathy on irony processing, correlating participants' empathic concern score on the IRI with the difference in P600 amplitude between the ironic and literal conditions. This difference was calculated by subtracting the average amplitude in the P600 time window in the literal condition from that in the ironic condition. Therefore, a positive value indicates a larger P600 effect for the irony condition. In the irony comprehension questionnaire, no statistical analysis was performed to compare age groups as all participants had an accuracy rate of 90% or above.

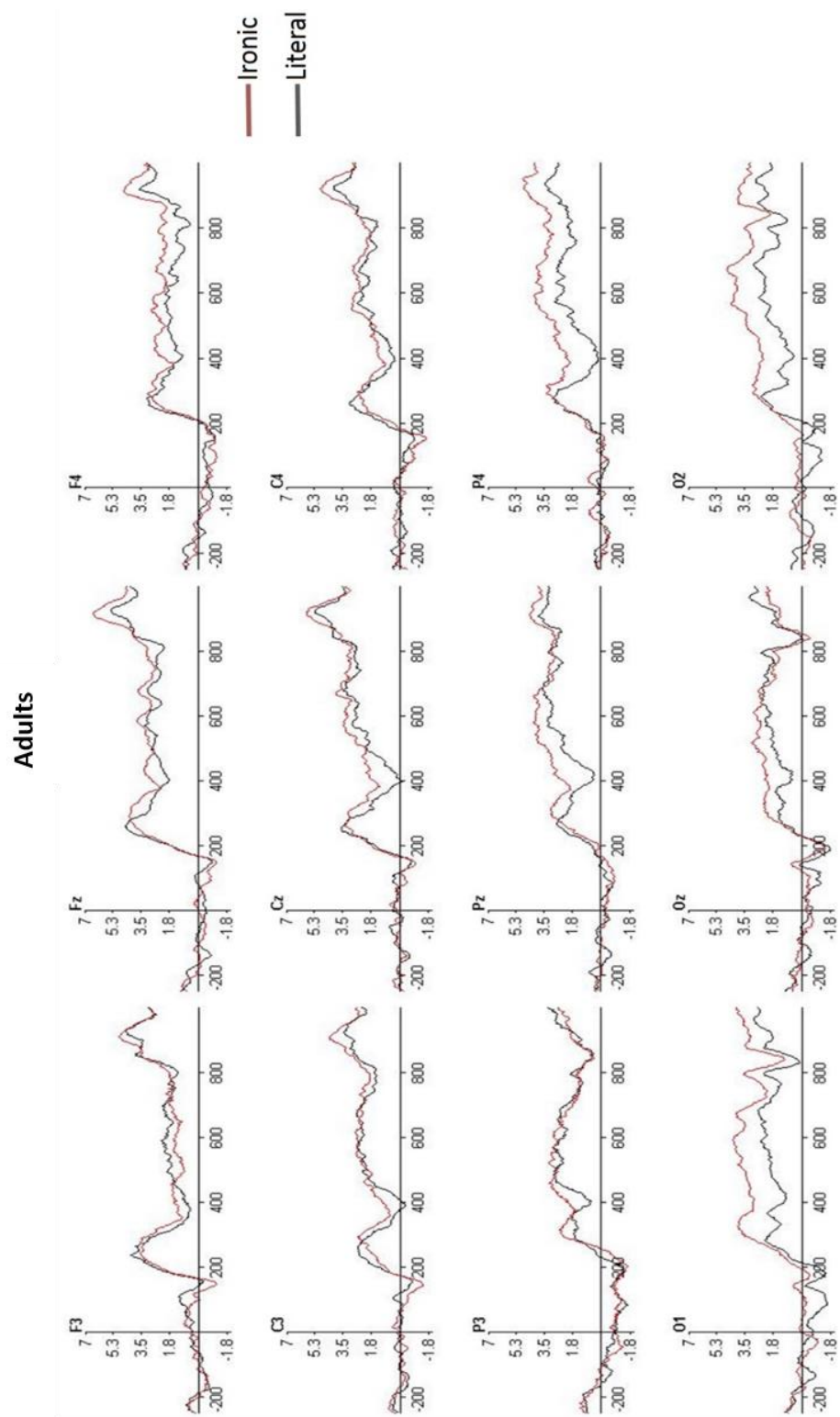
### 5.3 RESULTS

#### 5.3.1 ERP results

Grand average ERP waveforms of the 12 electrodes used in the analysis are shown in **Figure 14 (a-b)** for each age group. Scalp topographies of each age group are shown in **Figure 15**. It can be seen that both groups display a positivity. However, the effect in the adults is posterior, where as in the adolescents it is more broadly distributed.

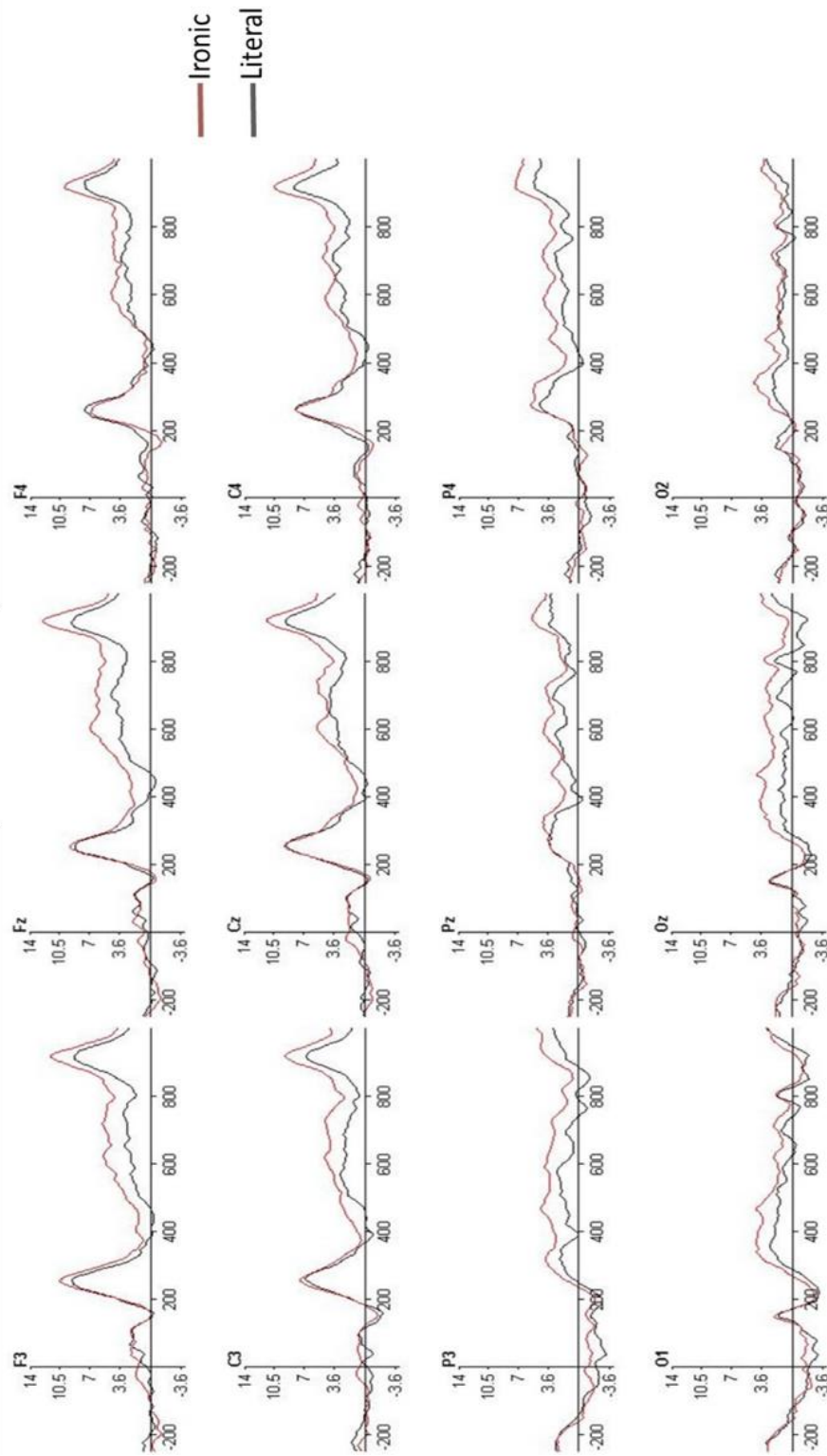
---

<sup>10</sup> Note that the 300-500ms window is not included in the analysis following Spotorno et al. (2013) (and previous ERP studies on irony), where no N400 effect was found in adults. However, given the additional adolescent group here the 300-500ms window was explored. Visual inspection of the waveforms suggests an early positivity rather than an N400. Exploratory analysis of the 300-500ms window suggests an effect of Story Type indicating a positivity ( $F(1,44) = 15.97$ ,  $p = .001$ ,  $\eta_p^2 = .266$ ), and therefore, we can be confident that there is no N400 effect in either age group.

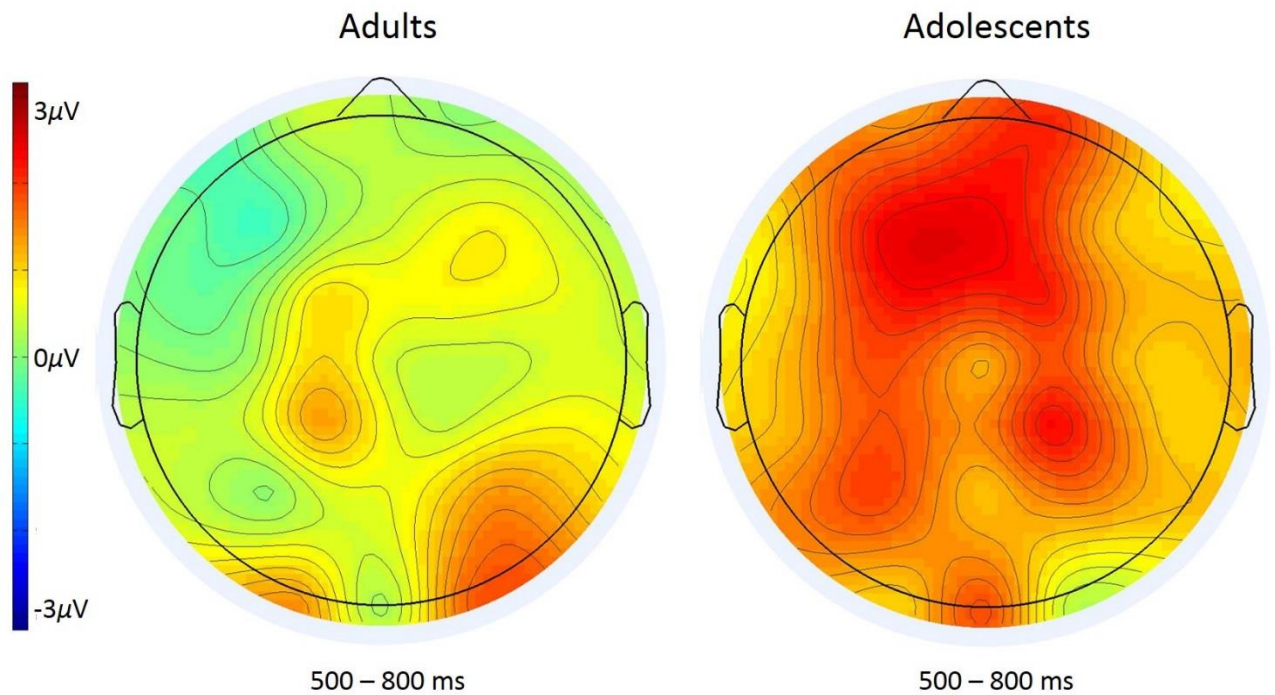


**Figure 14(a)** Grand average ERPs over twelve electrodes by the critical word in the target sentence for each Story type.

## Adolescents



**Figure 14(b)** Grand average ERPs over twelve electrodes by the critical word in the target sentence for each Story type.



**Figure 15** Topographic maps of the ERP difference waveform of the Story type (Ironic minus literal) for each age group between 500 and 800 ms from critical word onset.

**Table 17** shows the results of the main repeated measures ANOVA. Analyses showed no main effect of Age group ( $p = .76$ ) but there was a significant main effect of Story type ( $F(1,44) = 7.89, p = .007, \eta_p^2 = .152$ ) such that across electrodes and both age groups the ERPs were more positive in the ironic condition than the literal condition. Results also revealed a significant three-way Story type x Age group x Laterality interaction which was further explored by a Story type x Laterality ANOVA in adults and adolescents separately.

**Table 17** Omnibus repeated measures ANOVA

	df	F
Omnibus ANOVA		
Age group	1, 44	0.09
Story type	1, 44	7.89**
Laterality	2, 44	7.83**
Anteriority	1, 44	4.90*
Age group x Story type	1, 44	1.21
Age group x Laterality	2, 44	0.10
Age group x Anteriority	1, 44	13.64**
Story type x Laterality	2, 44	0.01
Story type x Anteriority	1, 44	0.07
Laterality x Anteriority	2, 44	6.86**
Story type x Age group x Laterality	2, 44	4.90*
Story type x Age group x Anteriority	1, 44	2.33
Story type x Laterality x Anteriority	2, 44	1.98
Age group x Laterality x Anteriority	2, 44	0.16
Age group x Story type x Laterality x Anteriority	2, 44	1.39

\*\* $p < .01$

\* $p < .05$

In adolescents there was a main effect of Story type ( $F(1,22) = 5.24$ ,  $p = .032$ ,  $\eta_p^2 = .192$ ), but no significant interaction between Story type and Laterality ( $p = .19$ ) suggesting that the positivity was broadly distributed. The adults however, showed a significant interaction of Story type x Laterality ( $F(2,22) = 4.22$ ,  $p = .021$ ,  $\eta_p^2 = .161$ ). Follow-up analysis of this interaction showed an effect of Story type in the right electrodes ( $F(1,22) = 6.68$ ,  $p = .017$ ,  $\eta_p^2 = .233$ ) but not in the left or midline electrodes (all  $ps > .27$ ). In sum, ironic stories elicited a greater positivity (a P600 effect) in both age groups; the effect was broadly distributed in adolescents but right-lateralised in adults.

### 5.3.2 Correlations with empathy

As mentioned above Pearson's correlations examined the relationship between participants' empathic concern score with the difference in amplitude between the



ironic condition and the literal condition. Given the topographic age group differences found in the ERP results, correlations were performed separately for each age group. For both age groups, correlations were performed with a by-subject average ERP difference at all twelve of the electrodes used in the analysis (F3, C3; Fz, Cz; F4, C4; P3, O1; Pz, Oz; P4, O2). In adult participants a significant positive correlation was found between empathy and the P600 effect [ $r(23) = 0.51, p=0.013$ ], suggesting that adults with a higher empathy score have a larger P600 effect during irony processing.<sup>11</sup> In contrast to adults, results showed no correlation between empathy and the size of the P600 effect [ $r(23) = -0.22, p=0.32$ ], suggesting that in adolescents individual differences in empathy do not seem to have an effect on irony processing.<sup>12</sup> Therefore, while individual differences in empathy seem to have an effect during irony processing in adults, it does not seem to be the case for adolescents.

## 5.4 DISCUSSION

The current experiment aimed to i) replicate Spotorno and colleagues' (2013) results from the adult participants, ii) investigate how adolescents' brains process irony in real time in comparison to adults and iii) examine whether individual differences in empathy may be associated with irony processing. Participants' ERPs were recorded while they read seven-sentence stories that only differed by two sentences. The target sentence was identical between conditions but it was either literal or ironic depending on the preceding context. Participants' individual differences in empathy were also measured through the IRI (Davis, 1983).

### 5.4.1 P600

The ERP analysis focused on the P600 component based on recent ERP studies that have consistently shown to be elicited during irony compared to literal language. Considering the adult data first, the critical word elicited a greater positivity in 500-

---

<sup>11</sup> Note that a significant correlation between empathy and the P600 effect is still found when performed with a by-subject average ERP difference at the electrodes on the right side of the scalp only (F4; C4; P4; O2) where results showed an effect of Story type [ $r(23) = 0.57, p=0.005$ ].

<sup>12</sup> Note that one-way ANOVA analysis showed that adolescents and adults do not differ in their empathy score ( $F(1,44) = 2.95, p = .093$ ).

800ms window in the ironic condition compared to the literal condition. Therefore, the P600 effect was replicated as seen previously in Spotorno and colleagues (2013), as well as other recent ERP studies (Regel & Gunter, 2016; Regel et al., 2011, 2014). Our observation of posterior positivity in adults is in line with previous irony comprehension ERP studies (Regel & Gunter, 2016; Regel et al., 2011, 2014). Interestingly, Spotorno and colleagues (2013) found a more frontal distribution than the current results and previous findings. This difference needs further investigation.

While adolescents also showed an enhanced P600 component to ironic critical words, there were topographic differences between adults' and adolescents' brain responses. Specifically, while in adults the effect was right-lateralised, it was broadly distributed in adolescents. The observed topographic difference between adults and adolescents could be due to structural or functional differences in the brain areas involved in ToM.

In terms of neuroanatomical changes, I reviewed in Chapter 1 the MRI research that has shown developmental changes in grey and white matter structures through adolescence into early adulthood. Specifically, studies have shown that while white matter volume increases through adolescence into early adulthood (Aubert-Broche et al., 2013; Barnea-Goraly et al., 2005; Giedd et al., 1999; Lebel & Beaulieu, 2011; Sowell et al., 1999), grey matter volume increases throughout childhood where it peaks and then declines across adolescence into adulthood (Aubert-Broche et al., 2013; Lebel & Beaulieu, 2011; Tamnes et al., 2013). As reviewed in Chapter 1, Mills and colleagues (2014) longitudinal neuroimaging study investigating the structural developmental trajectory of the brain areas involved in ToM (ATC, MPFC, TPJ, pSTS) found that grey matter volume and cortical thickness decreased in MPFC, TPJ and pSTS from childhood into the early twenties (Mills et al., 2014). In the ATC grey matter volume increased until around the age of 12 and then decreased and cortical thickness increased until early adulthood.

In addition to the anatomical changes, fMRI studies have shown changes in functional recruitment of the ToM network between adolescence and adulthood. As presented in Chapter 1, several studies using a variety of ToM tasks have consistently found MPFC activity decreases between adolescence and adulthood (Blakemore, 2008;

2012). The range of ToM tasks that have reported this finding include thinking about social emotions such as embarrassment (Burnett & Blakemore, 2009), thinking about one's own intentions to carry out an action (Blakemore et al., 2007), attributing emotional states of others (Gunther Moor et al., 2011), evaluating affective mental states of actors in video clips (Vetter, Altgassen, et al., 2013) and responding to cartoon scenarios that involve affective ToM (Sebastian et al., 2012). Crucially, as mentioned in the introduction, Wang and colleagues (2006) reported the same functional difference as in the other fMRI studies when processing ironic versus literal remarks. Given the evidence of the prolonged structural development of the ToM network, as well as changes in functional activity, the topographic differences that we observe in the current results could be a possible reflection of these changes in the brain during adolescence.

#### **5.4.2 Empathy and irony processing**

Added to the topographic differences, we found differences in the relation between empathy and irony comprehension between the two age groups. While adult participants showed a significant positive correlation between empathy and the P600 effect, no correlation was found in adolescents. This finding suggests that individual differences in empathy are associated with irony processing in adults but not in adolescents.

Shamay-Tsoory and colleagues' (2005; 2010) ToM model may provide an explanation for the lack of empathy correlation in adolescents. This view suggests a differentiation between cognitive ToM and affective ToM which involves cognitive ToM and empathy. According to Shamay-Tsoory and colleagues (2005), there are three stages of processing in irony comprehension. Firstly, the language areas in the left hemisphere are involved in interpreting the literal meaning of the utterance. Secondly, the intentional, social and emotional context is processed in the frontal lobes and the right hemisphere at which point the inconsistency between the literal meaning of the utterance and the social emotional context is detected. Crucially, the final stage involves the vMPFC, where the listener integrates the literal meaning of the utterance with the social and emotional information of the context and determines the intended meaning of the utterance. The authors suggest that this final stage of the process requires

affective ToM as one needs to have an empathic understanding of others' feelings to understand the speaker's attitude during irony comprehension. As stated previously, Shamay-Tsoory and colleagues (2010) proposed that affective ToM is the integration of cognitive ToM and empathy. The empathy correlation found in adults then is in line with this model.

Within the context of the above model, a possible explanation for the lack of association between adolescents' empathy and ERP response to irony might be that the ToM inferences involved here require integration of different processes such as empathy and adolescents struggle with this integration. Based on this model, a speculative hypothesis is that although adolescents engage cognitive ToM to recognize irony, they did not fully process speaker attitude. Although made with caution, this suggestion is in line with developmental behavioural data that showed that speaker intention is last to develop during middle childhood and into early adolescence (Demorest et al., 1984; Filippova et al., 2008; Pexman et al., 2005). Nicholson and colleagues (2013) suggested that the prolonged development of irony appreciation beyond middle childhood and into adolescence could be due to the fact that irony appreciation requires advanced skills to detect speaker's attitude and evaluate "the affective consequences" of their utterance. I suggest that these advanced ToM abilities are still developing during the adolescent years. Note that the suggestion here does not go as far as suggesting that adults use a different strategy to process irony. All that is proposed here is that these two age-groups process irony differently based on the topographical differences and the differences in empathy correlation. Based on this I put forward different possible avenues that can be explored further as to what might be different in adolescents.

This interpretation of the data is also in line with previous studies that found adolescents' performance is poor on affective ToM tasks but adequate on cognitive ToM tasks (Sebastian et al., 2012; Vetter, Altgassen, et al., 2013) which were discussed in Chapter 1 and 4. Moreover, findings from Chapter 4 suggest that when ToM demands are higher due to the involvement of emotional components and require complex integration online of mental state information adolescents perform more poorly than adults. These findings support the proposal that the differences observed here could be

due to adolescents' inability to use information about others' emotional mental state online as quickly as adults. It follows then that children's and adolescents' prolonged development of irony appreciation, specifically understanding speaker attitude, may be due to their difficulty in rapidly integrating more complex ToM inferences.

Based on the topographic differences and the differences in empathy association between the two groups, it seems then that adolescents are not processing irony the same way as adults. I acknowledge that both groups exhibited a P600 effect. However, the P600 effect in adolescents could simply reflect their detection of the inconsistency between the target sentence and the preceding context in the ironic condition, instead of them engaging ToM to comprehend irony in the same way as the adults. This hypothesis requires further investigation and I turn to outline a possible task to explore this hypothesis.

I take the current findings to suggest that the P600 effect observed in both age groups are functionally different due to the differences found in its topography as well as in its relation with empathy. One possible way of testing this hypothesis would be to compare participants' response to irony against materials that have been shown to elicit a P600 effect but that do not require ToM processes, such as semantic reversal anomaly (e.g. "That cat that fled from the mice" vs "The mice that fled from the cat", van Herten et al., 2005). This hypothesis would predict a P600 effect in both age groups for both types of materials. Crucially, I would expect the P600 effects to have similar topographical distribution across groups for the semantic reversal materials but not for the irony materials. Moreover, no associations between empathy and the P600 elicited in the semantic reversal anomaly task should be found in both age groups, while an association would be present only in adults during irony processing. These results would further support the proposal that adults and adolescents process irony differently. It would also provide further evidence for the proposal that these differences in irony processing could be due to the ongoing development of online use of ToM. I would further suggest that this follow up study would be done with more adolescent age groups that have a smaller age range (e.g. 11-13, 14-16 and 17-19). The age range in the current study was big (11-18). Having more age groups with a smaller age range can

disentangle whether these age-related differences are found only during early adolescence or if they continue through to late adolescence as well.

In addition to the proposed study, it should be noted that future analyses of the data using Time Frequency Analysis (TFA) could also shed light on the current results. In Spotorno and colleagues (2013) TFA analysis revealed an increase in power in the gamma band during the Ironic condition compared to the Literal one. Based on previous evidence that the gamma band can “index social dimensions of communication” (Van den Brink, Van Berkum, Bastiaansen, Tesink, Kos M, et al., 2010; Grossmann, Johnson, Lloyd-Fox, Blasi, Deligianni, et al., 2008), along with the evidence suggesting that the ToM network is involved in irony processing (Spotorno et al., 2011), the authors suggest that this increase might reflect social cognitive processes involved during irony processing (Spotorno et al., 2013, p. 8). Therefore, it would be fruitful to examine whether there is an increase in gamma power in adolescent participants as there is in adults. This could shed light on whether social cognitive processes are engaged in adolescents in the same manner as adults during irony processing.

#### **5.4.3 Contributions to the Irony processing debate**

Current debates on irony processing are concerned with the time-course of irony processing, specifically on the extent to which the literal interpretation is involved during irony comprehension. As the current study’s focus was on age-related differences in irony processing based on ongoing ToM development, the present results do not directly address questions about whether and when the literal interpretation is accessed during irony processing. However, it is worth pointing out briefly the main current accounts on irony processing and how the replicated adult data further contribute to the debate as previous studies that have found the same results (Regel et al., 2010, 2011, 2014; Spotorno et al., 2013).

Modular theories (Standard Pragmatic Model, Graded Salience Hypothesis) propose that the literal interpretation of an ironic utterance is accessed first before any non-literal interpretations are considered. One of the most prominent of these is the Standard Pragmatic Model (SPM), based on Grice’s (1975, 1989) theoretical account,

suggests that the literal meaning must be accessed first and then rejected as there is a mismatch between the literal meaning and the contextual information. As such, extra processing costs are required to derive richer interpretations that are contextually appropriate. The Graded Salience Hypothesis (Giora, 1997) another influential modular view, posits that upon processing an ironic utterance what is accessed first is the most salient meaning, regardless of whether it is the literal or the ironic. The salient meaning of the ironic utterance is the one that is encoded in the listener's mental lexicon. Therefore, whether the literal or the figurative meaning of an ironic remark will be accessed first, depends on which meaning is encoded in the lexicon. It is suggested that for ironies that are uncommon, more inferential processing and consequently more cognitive effort is required to derive the appropriate interpretation. According to modular views, deriving the ironic meaning requires extra inferential processes and therefore, should take longer to process than the equivalent literal meaning. As suggested by previous studies, an enhanced P600 for the ironic utterances in comparison to the literal ones suggests that the ironic utterances were more demanding and support the view that additional cognitive resources might be involved for the pragmatic interpretation (Regel et al., 2010, 2011, 2014; Spotorno et al., 2013).

In contrast to modular views, interactive approaches propose that ironic utterances do not need to be interpreted in a sequential manner. The Direct Access view suggests that lexical and contextual information are integrated early and therefore the figurative interpretation is accessed directly and the literal meaning is not considered (Gibbs, 1994, 2002). This view suggests then that no extra processing effort is required in processing an ironic remark in comparison to a literal one. However, as mentioned above, this does not explain why ironic utterances consistently elicit larger P600 responses than literal utterances.

A more recent interactive approach- the Parallel Constraint Satisfaction - suggests that multiple cues are activated and are considered rapidly in parallel (A. Katz, Blasko, & Kazamerski, 2004; A. N. Katz, 2005; Pexman, 2008). An ironic interpretation is then considered as soon as there is enough evidence from the activated cues to support it as the appropriate interpretation. Under certain circumstances, no extra processing step is needed and therefore, an ironic statement should not necessarily take longer to

process the literal equivalent. Spotorno and colleagues (2013) suggested that the constraint-satisfaction approach is in line with data showing extra processing in irony comprehension (i.e. the enhanced P600). Overall, the work here further supports accounts of irony comprehension that take into account the complexity and cognitive effort involved.

#### 5.4.4 Conclusion

To sum up, the current results showed that in both adolescents and adults, ironic stories elicited a greater positivity (a P600 effect) compared to the literal stories. However, there was a topographic difference between the P600 effects observed. The effect was broadly distributed in adolescents but more posterior in adults. Moreover, there was a significant positive correlation between empathy and the P600 effect in adults, but no correlation was found in adolescents. Based on these two differences, I propose that adolescents do not process irony in the same manner as adults do.

It was suggested that the topographic differences might be due to structural changes that are occurring in the brain during adolescence. Further, based on the differences found in the correlation between empathy and the P600 effect, I proposed a hypothesis that adolescents are not specialised irony comprehenders due to their ongoing development in ToM abilities. I suggested that adolescents are using their cognitive ToM to recognize irony but they are not adult-like in processing irony because they do not fully understand the speaker's intent, as proposed by other developmental studies. This suggestion is motivated by the lack of an empathy association in adolescents that was revealed in adults, as understanding speaker intent involves affective ToM processes which require a combination of empathy and cognitive ToM. However, as stated above this is only a hypothesis that requires further investigation with studies such as the one suggested above.

Considering the above interpretation of the data I also propose that the functional interpretation of the P600 in adolescents differs from that of adults for two main reasons. Firstly, the P600 has been implicated in many linguistic and non-linguistic phenomena and secondly, fMRI data show functional differences between adults and adolescents not only during a variety ToM tasks but also during irony comprehension.



Nevertheless, as this study was the first to investigate figurative language comprehension or ToM processes<sup>13</sup> in adolescents through ERP, further investigation is required to examine this proposal more closely.

---

<sup>13</sup> With the exception of one recent EEG paper that will be published at the end of this year, which has shown preliminary evidence of intact perspective taking in a narrative context in adolescents (10-15 year olds) (Westley et al., 2017). They found that semantically anomalous sentences elicited an N400 effect but that a social N400 effect was also elicited with sentences that are semantically plausible for the participants when they believe that the confederate find these sentences implausible. However, further research is needed to validate the effects as, i) there was no adult group to compare and, ii) there was no appropriate control condition where both the confederate and the participant read all the sentences.

## 6 CHAPTER 6: DISCUSSION

---

### 6.1 SUMMARY OF FINDINGS

The aim of the current thesis was to investigate the development of ToM in adolescence and what factors account for this age-related improvement. Specifically, I wanted to explore what psychological or cognitive processes might be developing in young and older adolescents in parallel to the brain changes that are occurring in the ToM network. As such, I tested adolescents aged 9-19 with four different online ToM tasks.

The first task, Chapter 2, used time-sensitive visual world techniques to follow up Dumontheil and colleagues (2010), which was the first and one of the main studies that found age-related differences in ToM use between adults and adolescents. Our behavioural results replicated the findings of Dumontheil and colleagues (2010) and were in fact more robust. Younger and older adolescents made more errors than adults only in the Director Experimental trials, which required taking the perspective of the Director into account (Director Experimental trials), in contrast to the rule-based control condition (No-Director Experimental trials). Added to that, the eye tracking data revealed an interesting pattern. When considering correct and incorrect trials separately, adults and adolescents showed the same eye-gaze pattern. Importantly, in correct and incorrect trials all age groups committed to a target at the earliest point in the linguistic stimulus, suggesting that the decision is made early on in the trial. All age groups seem to engage the same kind of online processes during perspective taking but differ in how often they do it. The question that was raised then is why does perspective taking fail more often in adolescents than adults? The results showed that IC was only a partial contributing factor. I suggested that given the subtle cues and complex integration processes involved in the task, what might still be maturing in adolescents is the ability to attend to and integrate relevant mental state information.

In Chapter 3 I further explored the above hypothesis by testing adults and adolescents with an online FB task which involved the same ToM complexity but eliminated constraining factors in adolescents' performance that I suggested in Chapter

2. The results showed that when available cues are salient and EF demands are lowered, adolescents can infer other people's perspective during false-belief reasoning as spontaneously as adults. These results support the hypothesis that attending to and integrating relevant ToM information rapidly from subtle cues is an ability that is still in development during adolescence. However, these results also called for further investigation as to whether other factors play a role over and above processing efficiency.

The tasks in Chapters 2 and 3 both required participants to make an inference based on a first-order belief of another agent which differed from her/his belief and, therefore, these tasks did not differ from one another in terms of the inferential complexity involved. In Chapter 4 I investigated whether adolescents and adults might differ in a task that involves a greater degree of ToM inferential complexity but that involves comparable EF demands and salience of relevant mental state information to the FB task in Chapter 3. Specifically, I tested whether adolescents can use knowledge about a character's preferences and higher-order desires to make complex ToM inferences and predict that character's subsequent behaviour as quickly as adults. Despite, eliminating possible factors that are argued to interfere (i.e. conflict from linguistic input or knowledge from the self-perspective) and keeping EF demands low, in comparison to adults, all adolescents suffered a delay in generating expectations based on mental state information. Note that the suggestion here is not that adolescents are only just acquiring the ability to do second order ToM, but that this online task showed that adolescents and adults differ in how rapidly they integrate the relevant inferences. These results suggest that WM and IC are not the only contributing factors in the application of ToM. Note also that, unlike in the DT, the relevant mental state information is highly salient, as it is provided in the context sentence. Given that EF demands were low in the current task and that mental state information is salient and available to integrate, I suggested that what adolescents struggle with here is the ToM inferential complexity involved. I further proposed that what could be more 'complex' for adolescents is that the inferences one has to make to predict behaviour are based on higher-order mental states. In addition, there are emotional components they have to consider and integrate. Further investigation would benefit to disentangle whether

one of these is a constraining factor for adolescents or whether it is the combination of both.

Finally, in Chapter 5 I further explored adolescents' complex ToM inferential abilities by investigating the time-course of irony comprehension through ERPs, in comparison to adults. While in both age-groups, ironic stories elicited a greater positivity (a P600 effect) compared to the literal stories, the effect was broadly distributed in adolescents but more posterior in adults. Moreover, there was a significant positive correlation between empathy and the P600 effect in adults, but no correlation was found in adolescents. Although topographic differences might be due to structural changes that are occurring in the brain during adolescence, I argued that this does not explain the differences found in the correlation between empathy and the P600 effect. Considering the evidence found in Chapter 4 as well as the results in Chapter 5, I proposed that adolescents' processing of irony differs from adults due to their ongoing development of abilities I suggested above. Specifically, I posited that adolescents recognize irony perhaps through cognitive ToM abilities, but they are not adult-like in processing irony because they do not fully understand the speaker's intent. This suggestion is driven by the lack of an empathy association in adolescents but that was present in adults, as understanding speaker intent involves affective ToM processes, which require a combination of empathy and cognitive ToM. This is one of the first studies to investigate irony comprehension (or any figurative language comprehension) and ToM processes in adolescents through ERPs. Therefore, this study is only a starting point that I believe warrants further investigation.

## **6.2 THE BIG PICTURE**

In Chapter 1 I reviewed the substantial number of studies that have shown how the social brain undergoes changes during adolescence. There are a few behavioural studies that have found some evidence of age-related differences between adolescents and adults and even less so in older adolescents. It was suggested that functional differences found between adults and adolescents might reflect a difference in cognitive strategy (Blakemore, 2008; Dumontheil, 2015). One proposal is that adults retrieve 'social scripts' based on previous experiences and adolescents process new

computations to interpret the same social situation as they are less experienced. However, beyond this hypothesis, the current thesis makes a theoretical advance as to what changes cognitively (or psychologically) in ToM during adolescence.

Children are at ceiling in FB tasks by the age of 7 so we know that they possess ToM by the time they reach adolescence. Therefore, the question arises as to what could develop in terms of cognitive (psychological) processes? Outside of ToM processes, there is evidence that EF develops across adolescence and we have seen from Chapter 2 and previous work (Vetter, Altgassen, et al., 2013) that IC does have an impact. However, Chapters 2 and 3 together provide evidence that EF development is not the only factor as by manipulating the salience of the mental state information adolescents perform as well as adults.

Taken together, I propose that Chapters 2-4 suggest two possible abilities that are still developing. First, although adolescents can integrate mental state information as rapidly as adults when that information is salient and task demands are otherwise low, adolescents do not seem to be able to detect mental state information and integrate it into online processes as well as adults. Second, although adolescents 'have' ToM, they do not seem to integrate more complex ToM inferences. As I suggested in Chapter 4, the 'complexity' here for adolescents might be the fact that inferences are based on higher-order mental state or because there are emotional components they have to consider and integrate. Chapter 5 further supports this second hypothesis as we found differences on how adolescents and adults process irony comprehension which requires complex inferences about a speaker's beliefs and negative or mocking attitude and associated emotions. Overall, therefore, the findings from the current thesis provide further evidence that ToM is still developing during adolescence but crucially, advances our knowledge on what specific cognitive processes might still be developing beyond middle childhood and into adolescence.

Some of the findings in the current thesis can be explained through the social scripts proposal. It could be that in a given social situation adults are better able to attend to and integrate available social information from surrounding stimuli rapidly because they retrieve these social scripts. Whereas adolescents do not have enough experience to detect the relevant information and, therefore, might fail altogether in

interpreting that same social situation or they have to process new computations which takes time. However, social scripts require flexible application given a specific situation and actual information in the context. The results in Chapter 2 suggest that adolescents have the specific social script because they are as quick as adults are when they are successful, but they are not as successful perhaps because they do not always apply the script using the right information in the context. The findings in Chapter 4, however, suggest that perhaps adolescents do not yet have a social script for higher order desires or they are slower at combining highly salient information during the task.

### **6.3 IMPLICATIONS FOR THEORIES ON THE ROLE PERSPECTIVE INFORMATION HAS DURING LANGUAGE PROCESSING**

As discussed in Chapter 2 there is considerable debate in language processing literature about the role of ToM – how it might be involved and when - in online language processing (Brown-Schmidt & Hanna, 2011). As presented in previous chapters, Constraint-based models posit that perspective information can have an effect early on in language processing and that the extent to which mental perspective information is integrated on-line depends on the extent to which i) other constraints are conflicting, ii) how salient or available the perspective information is (Brown-Schmidt & Hanna, 2011; Samson et al., 2010), iii) how sensitive a participant is to such information (Brown-Schmidt, 2012) or their ability to integrate that information with linguistic and visual contextual information (Hanna et al., 2003). Conversely, the Perspective-Adjustment model (the egocentric account) posits that listeners are initially biased to associate information to their own egocentric perspective and that successful use of knowledge of someone else's belief only occurs as a (costly) second process which corrects any manifest errors (Keysar et al., 2000, 2003). The findings from this thesis (Chapter 2-4) contradict the Perspective-Adjustment model (the egocentric account) which I discuss below.

In Chapter 2 I argued how the results are not in line with the Perspective-Adjustment model. Based on the egocentric account, participants on incorrect trials during the Director task should first attend to the distractor which would be the best-fit referent if they apply an egocentric bias, and then remain focused on the distractor as

the second adjustment process would fail. However, results revealed that participants considered the target early on in the trial and their eye-gaze only shifted to the distractor right before they responded. Therefore, I argued that eye-gaze data showed an influence of the speaker's perspective early on which is in line with Constraint-based models (Hanna et al., 2003; Brown-schmidt, Gunlogson, & Tanenhaus, 2008; Heller et al., 2008; Brown-Schmidt & Hanna, 2011). This hypothesis, however, could benefit from follow up studies that focus on adult participants.

In Chapter 3 I considered whether perhaps the delay both age groups showed in predicting the appropriate container during FB trials, in comparison to TB trials could be explained from the egocentric account. According to the Perspective-Adjustment model (Epley et al., 2004; Keysar et al., 2000, 2003) this delay would be explained a default initial bias to take an egocentric perspective which is later adjusted, during a second process to the speaker's perspective. However, I argued that this does not explain why listeners show significantly different visual biases between TB and FB trials from when they hear whose perspective to take, showing a clear early sensitivity to the different perspectives in both conditions. I further argued that if egocentric biases are the default in ToM processing, then a difference in visual biases should not be occurring early on in the trial. Instead divergence would occur at later stage when listeners would adjust their perspective according to the character's knowledge. Moreover, as I mentioned in Chapter 3, results from the active group in Ferguson and colleagues (2015) also contradict this view as participants correctly predict the initial location in FB trials from when they hear whose perspective to consider.

I also examined whether the delay adolescents suffered in Chapter 4 could be explained by the Perspective-Adjustment model. By this account, adolescents have a larger egocentric bias and it is harder for them to correct that egocentric propensity. However, this view is inconsistent with the adult data here, as well as the results from the original Ferguson & Breheny (2011) study, that show that participants can predict the character's action quickly and accurately based on the character's preferences and intentions, even when those desires are in conflict. The results suggest that they are able to use information about others' mental states very early on and predict the character's action. Further studies that have challenged this view have

shown that people spontaneously infer other people's' perspectives, even when there is no explicit reason to do so (Ferguson, Apperly, et al., 2015; Samson et al., 2010; Schneider, Nott, & Dux, 2014), as we have also found in Chapter 3.

## **6.4 WIDER IMPLICATIONS**

While the findings in the current thesis provide specific insights on how adolescents develop the ability to understand other people, there are some wider implications that I believe are worth mentioning.

### **6.4.1 Autism**

One of the prominent cognitive abnormalities that has been widely acknowledged in autism spectrum disorder (ASD) is ToM (Brunsdon & Happé, 2014; Frith, 2012). It has been argued that limited ToM in individuals with ASD can be viewed as an egocentric bias – a propensity to overestimate the similarity of other people's experiences compared to their own (Begeer et al., 2016; Frith & de Vignemont, 2005; Goldman & Sebanz, 2005). As reviewed in Chapter 3 a recent study tested typically developing (TD) 7-13 year olds as well as children with ASD using a modified false belief task- the Sandbox task- and a visual hindsight bias task (Begeer et al., 2016; see Chapter 3 for the details of the study). The authors argue that the Sandbox allows for a reliable measure of egocentric bias as it has been shown in typically developing children, young adults (Sommerville, Bernstein, & Meltzoff, 2013) and older adults (Bernstein, Thornton, & Sommerville, 2011). Surprisingly, they found no differences between the two groups (with and without ASD). Children and young adolescents with and without ASD showed similar egocentric tendencies. The authors argue that egocentric bias in TD has been underestimated and perhaps has led to ASD children appearing more deficient in ToM reasoning. The current thesis provides some evidence of what might still be developing in adolescents. Having a better understanding of the social cognitive processes developing in typically developing adolescents can benefit research in ASD adolescents to better identify differences between the two groups.



#### **6.4.2 Implications for society**

One of the bigger questions is whether this ongoing development can have any impact on adolescents' well-being, be it directly or indirectly. At the beginning of my thesis I stated that adolescents face a lot of significant changes but one thing that certainly becomes more complex is their social world. Compared to children, adolescents begin to become more independent in choosing the environmental stimuli they experience which often leads to exposure to novel situations. Crucially though, they begin to form new peer relationships that are more complex and hierarchical (Brown, 2004), and these peer relationships become 'the centre of their world' as they transition into adolescence. ToM is a skill that is vital in navigating the emerging complexities of their social environment (Blakemore & Mills, 2014; Kilford et al., 2016) and crucially, their peer relationships. A reasonable assumption to make is that the ability to understand other's beliefs, desires, intentions and emotions in order to predict other's behaviour and understand what other's think if we can enable better peer relationships (Slaughter, Imuta, Peterson, & Henry, 2015). Indeed, ToM development has been associated with peer relationships in children and young adolescents (Banerjee, Watling, & Caputi, 2011 in 7-11 year olds; Bosacki & Astington, 1999 in 9-11 year olds; Caputi, Lecce, Pagnin, & Banerjee, 2012 in 5-7 year olds; Devine & Hughes, 2013 in 8-13 year olds; Slaughter et al., 2015 in 2-10 year olds). A meta-analysis (Slaughter et al., 2015) showed that children's (2-10 years old) ToM understanding is significantly related with peer group popularity. Based on these findings, the authors propose that children's ToM abilities can have an impact on their social environment with peers. Similarly, Devine & Hughes (2013) found that children and young adolescents who performed worse on the ToM tasks were more likely to report higher levels of peer rejection and loneliness (see Chapter 1 and 3 for a review of the study).

For adolescents, how they are viewed by their peers is of great concern to them (O'Brien & Bierman, 1988) and social exclusion can lower their mood (Sebastian, Viding, Williams, & Blakemore, 2010), and increase anxiety and social distress (Sebastian et al., 2011, 2010). In recent study adolescents (11-16 years old) and adults (22-47 years old) took part in online game called Cyberball, in which in certain conditions they were excluded from the game. After being excluded, in comparison to adults, all adolescents

reported a lowered mood and young adolescents (11-14 years) anxiety increased, suggesting that their ability to regulate distress after being socially excluded is still maturing (Sebastian et al., 2010). Social exclusion has been associated with other negative effects such as depression (Graham, Bellmore, & Juvonen, 2003; Isaacs, Card, & Hodges, 2001; Rigby, 2000, 2003), as well as long-term mental and physical health outcomes (Lev-Wiesel, Nuttman-Shwartz, & Sternberg, 2006; Prinstein, Sheah, & Guyer, 2005; Rigby, 2000). Moreover, fear of social exclusion has also been related to higher behavioural risk-taking in adolescents (Falk et al., 2014; Peake, Dishion, Stormshak, Moore, & Pfeifer, 2013). In another Cyberball game study (Peake et al., 2013) adolescents played a car-driving simulation, once before they Cyberball game, and once after whilst supposedly being observed by two peers over the internet. In the driving session after the Cyberball game where they were excluded, adolescents showed an increase in behavioural risk-taking (Peake et al., 2013). Along with previous findings, the current thesis suggests that adolescence is a period of life when skills necessary to navigate in a social environment are refined. If ToM abilities are viewed as important for children's social lives with peers, then they should be viewed just as crucial, if not more, during the adolescent years, considering the increased complexity of peer relationships and their hypersensitivity to peer rejection. I suggest that perhaps this protracted development in ToM can have an indirect impact on adolescents' well-being and should be further investigated.

While social functioning plays a role in adolescents' lives outside the classroom, it can also play a role inside the classroom in shaping learning and academic performance (Blakemore, 2008, 2012; Dumontheil, 2015). It has been suggested that it might be beneficial for education platforms to include training for abilities that are controlled by the brain areas that change the most during adolescence such as internal control, planning and ToM (Blakemore, 2012; Dumontheil, 2015). While it is easier to see how training for processes like planning and multitasking can be designed, it is harder for complex social cognitive processes such as ToM. However, a recent study tested a conversation based training program in 9-10 year olds and found that ToM abilities improved in the intervention group and this effect lasted over 2 months (Serena Lecce et al., 2014). The authors suggest that the training program enabled children to

be more sensitive and aware of when and how to use inferential information. Moreover, they propose that children became better at being able to detect relevant mental state information in certain social situations (Serena Lecce et al., 2014). There is also some evidence to suggest that reading certain types of literary fiction can strengthen ToM performance (Kidd & Castano, 2013; Kidd, Ongis, & Castano, 2016). Therefore, it might be fruitful to emphasize literary fiction in schools. However, in the United States for example, they recently adopted the Common Core State Standards which effectively lowers the importance of literary fiction in secondary education.

#### 6.4.3 Age of testing

Turning to a more procedural implication, the work in this thesis I believe has highlighted the issue of how old adult participants should be. The social brain undergoes such significant changes until the early twenties, and we now have robust evidence that psychological processes are developing in parallel during adolescence. Therefore, participants between the ages of 18-22 for social cognition studies would not necessarily represent adult performance. I raise this issue because most behavioural research in cognitive psychology and pragmatics use university undergraduate students as subjects which tend to be aged 18-22. Given that these changes occur until the early twenties, a more appropriate adult age group to test would be 24-35 years of age. During my research I found some evidence to support this concern when I compared the results from the adult participants in Chapter 4 who were 24-36 years old and the adult participants in the original study (Ferguson & Breheny, 2011) who were university undergraduate students. What is noticeable is that the adults in Ferguson & Breheny (2011) performed better than the adolescent group in Chapter 4 but they did not anticipate as early and robustly as the adult group in Chapter 4. Adults in the current study made significantly different visual biases between the Open and Secret conditions from as early as the ambiguous term (e.g. "car"), but not until the post-ambiguous region in the original study (Ferguson & Breheny, 2011). Note that in Chapter 2, the adult participants that were used were 19 to 29 year-olds in order to match the ages of the adult group used in the original Dumontheil and colleagues (2010) study since we aimed to replicate the study.

Another issue that has been previously raised about age groups when testing adolescents is that where possible (i.e. paradigm permitting), studies should include children, as well as adults. Fuhrmann, Knoll & Blakemore, (2015), proposed that studies should compare children, adolescents and adults to be able to investigate whether adolescence is a “stand-alone period of heightened plasticity, a continuous sensitive period with childhood, or if it represents a sensitive period at all.” (Fuhrmann et al., 2015, pp. 559–560). They suggest that if adolescence is a sensitive period specific stimulus should have a higher effect on brain and behaviour before or after, i.e. childhood or adulthood (Fuhrmann et al., 2015). The results from the current thesis support this proposal to some extent as we have seen that one task (Chapter 2) elicited age-related differences between older children and adolescents, where in another (Chapter 4) it did not. Note that the paradigms in Chapters 3 and 5 were not suitable for participants younger than 11, however, adult groups were always present for appropriate comparisons. Recent studies that have begun to investigate adolescents often do not even compare their findings to adult participants (Begeer et al., 2016; Devine & Hughes, 2013; Westley, Kohút, & Rueschemeyer, 2017), or they include young adults (Altgassen, Vetter, Phillips, Akgün, & Kliegel, 2014; Vetter, Leipold, et al., 2013; Vetter, Altgassen, et al., 2013), while others do not include older adolescents (14 and older) (Devine & Hughes, 2013; Pfeifer, Masten, Borofsky, Dapretto, & Lieberman, 2011; Wang et al., 2006).

## **6.5 FUTURE DIRECTIONS**

While the current research has shed some light on our understanding of what social cognitive processes are developing during adolescence, many questions remain to be investigated. Based on the findings from the current thesis, I highlight some possible directions that I think would further develop our understanding of social cognition development during adolescents.

A question that was raised from Chapter 4 was whether ToM abilities in adolescents are constrained when emotion is involved either because it is harder for them to recognize emotional states or integrating how the affective state will impact someone's desire. While there is some evidence that adolescents perform worse in

affective ToM, there have been no investigations as what processes specifically are a constraining factor for adolescents. What aspects of understanding and interpreting other emotions are still developing? Are there specific mechanisms that are still in development that constrain affective ToM? For example, there is evidence to suggest that face-processing abilities continue to develop during adolescents (Monk et al., 2003; Pfeifer et al., 2011) as well as emotional face recognition (Thomas, De Bellis, Graham, & LaBar, 2007). Fuhrman and colleagues (2016) found that face memory and face perception abilities mature between early and late adolescence. Facial expressions can be social cues and provide information about other's mental state and emotions (Fuhrmann et al., 2016). Does this ongoing development in face processing play a factor in ToM use during adolescents? Further research is required to explore adolescents' understanding of emotions and how that is involved in ToM processing.

Given the current evidence, a further question that is raised is to what extent this prolonged development has impact, if any, on figurative language processing and pragmatic phenomena in general. As it was previously thought that ToM abilities are fully intact by the age of 5, research on has focused on how early children can perform on different pragmatic phenomena such as implicatures (Chierchia, Guasti, Gualmini, Meroni, & Crain, 2004; Guasti et al., 2005; Katsos & Bishop, 2011; Noveck, 2001; Papafragou & Musolino, 2003; Pouscoulous, Noveck, Politzer, & Bastide, 2007), and metaphors (Pearson, 1990; Waggoner & Palermo, 1989). The results from the current thesis suggest that perhaps figurative language processing should be re-examined in adolescents given that certain abilities required to make pragmatic inference are late-maturing. In the current thesis this question has only been touched upon by investigating irony processing. However, the results warrant further investigation not only in irony processing but other figurative language processing, such as hyperbole and meiosis.

Finally, it would be beneficial for research to focus on how this protracted development in ToM affects real-life situations. How do ToM abilities enable adolescents to navigate within these complex social environments? For example, as suggested above, peer rejection seems to be associated with ToM abilities, but it is not yet clear whether adolescents who are better at ToM are better at peer relationships or whether adolescents are worse at ToM because they do not have as much exposure

with peers. It has also been suggested that social functioning plays a role in “shaping learning and academic performance” (Dumontheil, 2015, p. 122) and while there is some evidence that ToM performance in children is associated with school achievement (Lecce, Caputi, & Pagnin, 2014), it would be fruitful to explore this in adolescents.

## **6.6 CONCLUSION**

To conclude, the aim of the current thesis was to investigate ToM development in adolescents and to investigate what cognitive processes are changing in parallel to the Social Brain changes that are occurring during this period of life. Overall, the findings from the current thesis point to specific social cognitive processes that underpin adolescents’ progress in understanding other’s minds. Moreover, the current findings contribute to the overall question of how the structural and functional changes in the human brain during adolescence relate to social cognitive development. While the current thesis provided new insights into social cognition development during adolescents, it has also paved the path for future directions in the rapidly expanding field.

## REFERENCES

---

- Ackerman, B. P. (1981). Young children's understanding of a speaker's intentional use of a false utterance. *Developmental Psychology*, 17, 472–480.
- Ackerman, B. P. (1983). Form and Function in Children's Understanding of Ironic Utterances. *Journal of Experimental Child Psychology*, 35, 487–508.
- Altgassen, M., Vetter, N. C., Phillips, L. H., Akgün, C., & Kliegel, M. (2014). Theory of mind and switching predict prospective memory performance in adolescents. *Journal of Experimental Child Psychology*, 127, 163–175. <http://doi.org/10.1016/j.jecp.2014.03.009>
- Altmann, G. T., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, 111, 55–71.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264.
- Andrews, J., Rosenblatt, E., Malkus, U., Gardner, H., & Winner, E. (1986). Children's abilities to distinguish metaphoric and ironic utterances from mistakes and lies. *Communication & Cognition*, 19, 281–298.
- Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential theory of mind task. *Cognition*, 106, 1093–1108. <http://doi.org/10.1016/j.cognition.2007.05.005>
- Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is Belief Reasoning Automatic? *Psychological Science*, 17, 841–844. <http://doi.org/10.1037/0022-3514.87.3.327>
- Aristotle. (1905). *Aristotle's Politics*. Oxford: Clarendon Press
- Aubert-Broche, B., Fonov, V. S., García-lorenzo, D., Mouiha, A., Guizard, N., & Coupé, P. (2013). A new method for structural volume analysis of longitudinal brain MRI data and its application in studying the growth trajectories of anatomical brain structures in childhood. *NeuroImage*, 82, 393–402. <http://doi.org/10.1016/j.neuroimage.2013.05.065>
- Back, E., & Apperly, I. a. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition*, 115(1), 54–70. <http://doi.org/10.1016/j.cognition.2009.11.008>
- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Science*, 14(3), 110–118. <http://doi.org/10.1016/j.tics.2009.12.006>.False-belief
- Banerjee, R. (2000). The development of an understanding of modesty. *British Journal of Developmental Psychology*, 18, 499–517.
- Banerjee, R., Watling, D., & Caputi, M. (2011). Peer relations and the understanding of faux pas: Longitudinal evidence for bidirectional associations. *Child Development*, 82, 1887–1905. <http://doi.org/10.1111/J.1467-8624.2011.01669.X>
- Barnea-Goraly, N., Menon, V., Eckert, M., Tamm, L., Bammer, R., Karchemskiy, A., ... Reiss, A. . (2005). White matter development during childhood and adolescence: a cross-sectional diffusion tensor imaging study. *Cerebral Cortex*, 15, 1848–1854.
- Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux

pas by normally developing children with Asperger syndrome or high functioning autism. *Journal of Autism and Developmental Disorders*, 29, 407–418.

- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism Spectrum Quotient: Evidence from Asperger syndrome/high functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31(1), 5–17. <http://doi.org/10.1023/A:1005653411471>
- Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, 109(1), 18–40. <http://doi.org/10.1016/j.cognition.2008.07.005>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. (2013). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. <http://doi.org/doi:10.18637/jss.v067.i01>
- Begeer, S., Bernstein, D. M., Aßfalg, A., Azdad, H., Glasbergen, T., Wierda, M., & Koot, H. M. (2016). Journal of Experimental Child Reprint of : Equal egocentric bias in school-aged children with and without autism spectrum disorders q. *Journal of Experimental Child Psychology*, 149(April), 134–145. <http://doi.org/10.1016/j.jecp.2016.05.017>
- Bernhardt, B. ., & Singer, T. (2012). The neural basis of empathy. *Annual Review of Neuroscience*, 35, 1–23.
- Bernstein, D. M., Thornton, W. L., & Sommerville, J. A. (2011). Theory of mind through the ages: Older and middle-aged adults exhibit more errors than do younger adults on a continuous false belief task. *Experimental Aging Research*, 37, 481–502.
- Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, 18, 382–386.
- Blakemore, S.-J. (2008a). The social brain in adolescence. *Nature Reviews Neuroscience*, 9(4), 267–277. <http://doi.org/10.1038/nrn2353>
- Blakemore, S.-J. (2008b). The social brain in adolescence. *Nature Reviews Neuroscience*, 9(4), 267–277. <http://doi.org/10.1038/nrn2353>
- Blakemore, S.-J. (2012). Development of the social brain in adolescence. *Journal of the Royal Society of Medicine*, 105, 111–116.
- Blakemore, S.-J., den Ouden, H., Choudhury, S., & Frith, C. (2007). Adolescent development of the neural circuitry for thinking about intentions. *Social Cognitive and Affective Neuroscience*, 2(2), 130–9. <http://doi.org/10.1093/scan/nsm009>
- Blakemore, S.-J., & Mills, K. L. (2014). Is Adolescence a Sensitive Period for Sociocultural Processing ? *Annual Review of Psychology*, 65, 187–207. <http://doi.org/10.1146/annurev-psych-010213-115202>
- Bosacki, S., & Astington, J. W. (1999). Theory of Mind in Preadolescence: Relations Between Social Understanding and Social Competence. *Social Development*, 8, 237–255.
- Brown-Schmidt, S. (2009a). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, 61(2), 171–190. <http://doi.org/10.1016/j.jml.2009.04.003>



- Brown-Schmidt, S. (2009b). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, 16(5), 893–900. <http://doi.org/10.3758/PBR.16.5.893>
- Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes*, 27(1), 62–89. <http://doi.org/10.1080/01690965.2010.543363>
- Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation, 107(3), 1122–1134.
- Brown-Schmidt, S., & Hanna, J. (2011). Talking in another persons shoes: Incremental perspective-taking in language processing. *Dialogue & Discourse*, 2(1), 11–33. <http://doi.org/10.5087/dad.2011.102>
- Brown, B. B. (2004). Adolescents' Relationships with Peers. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of Adolescent Psychology* (Second Edi, pp. 363–394). Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Brunsdon, V. E., & Happé, F. (2014). Exploring the “fractionation” of autism at the cognitive level. *Autism*, 18(1), 17–30. <http://doi.org/10.1177/1362361313499456>
- Bull, R., Phillips, L. H., & Conway, C. A. (2008). The role of control functions in mentalizing: Dual-task studies of theory of mind and executive function. *Cognition*, 107, 663–672.
- Burnett, S., & Blakemore, S. (2009). Functional connectivity during a social emotion task in adolescents and in adults. *European Journal of Neuroscience*, 29, 1294–1301.
- Burnett, S., Sebastian, C., Cohen Kadosh, K., & Blakemore, S.-J. (2011). The social brain in adolescence: evidence from functional magnetic resonance imaging and behavioural studies. *Neuroscience and Biobehavioral Reviews*, 35(8), 1654–64. <http://doi.org/10.1016/j.neubiorev.2010.10.011>
- Cane, J., Ferguson, H. J., & Apperly, I. A. (2017). Using perspective to resolve reference: the impact of cognitive load and motivation. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, 43, 591–610. <http://doi.org/doi:10.1037/xlm0000345>
- Caputi, M., Lecce, S., Pagnin, A., & Banerjee, R. (2012). Longitudinal effects of theory of mind on later peer relations: The role of prosocial behavior. *Developmental Psychology*, 48, 257–270. <http://doi.org/10.1037/A0025402>
- Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, 87, 299–319. <http://doi.org/10.1016/j.jecp.2004.01.002>
- Channon, S., & Crawford, S. (2000). The effects of anterior lesions on performance on a story comprehension test: Left anterior impairment on a theory of mind-type task. *Neuropsychologia*, 38, 1006–1017.
- Chierchia, G. T., Guasti, M. T., Gualmini, A., Meroni, L., & Crain, S. (2004). Semantic and Pragmatic Competence in Children's and Adults' Comprehension of Or. In I. A. Noveck & D. Sperber (Eds.), *Experimental Pragmatics* (pp. 283–300). New York: Pelgrave Macmillan Ltd.
- Cohen, A. S., & German, T. C. (2009). Encoding of others' beliefs without overt instruction.

*Cognition*, 111, 356–363.

- Conklin, H. M., Luciana, M., Hooper, C. J., & Yarger, R. S. (2007). Working memory performance in typically developing children and adolescents: behavioral evidence of protracted frontal lobe development. *Developmental Neuropsychology*, 31(1), 103–28. <http://doi.org/10.1080/87565640709336889>
- Converse, B. a, Lin, S., Keysar, B., & Epley, N. (2008). In the mood to get over yourself: mood affects theory-of-mind use. *Emotion (Washington, D.C.)*, 8(5), 725–30. <http://doi.org/10.1037/a0013283>
- Coulson, S., & Kutas, M. (2001). Getting it: human event-related brain response to jokes in good and poor comprehenders. *Neuroscience Letters*, 316(2), 71–74.
- Coulson, S., & Lovett, C. (2010). Comprehension of non-conventional indirect requests: an event-related brain potential study. *Italian Journal of Linguistics*, 22, 107–124.
- Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: an event-related potential study. *Memory & Cognition*, 30(6), 958–968.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113–126. <http://doi.org/10.1037/0022-3514.44.1.113>
- De Grauwe, S., Swain, A., Holcomb, P. J., Ditman, T., & Kuperberg, G. R. (2010). Electrophysiological insights into the processing of nominal metaphors. *Neuropsychologia*, 48, 1965–1984. <http://doi.org/10.1016/j.neuropsychologia.2010.03.017>
- de Groot, A., Kaplan, J., Rosenblatt, E., Dews, S., & Winner, E. (1995). Understanding versus discriminating nonliteral utterances: Evidence for a dissociation. *Metaphor & Symbolic Activity*, 10, 255–273.
- Demorest, A., Meyer, C., Phelps, E., Gardner, H., & Winner, E. (1984). Words speak louder than actions: Understanding deliberately false remarks. *Child Development*, 55, 1527–1534.
- Demorest, A., Silberstein, L., Gardner, H., & Winner, E. (1983). Telling it as it isn't : Children's understanding of figurative language. *British Journal of Developmental Psychology*, 1, 121–134.
- Devine, R. T., & Hughes, C. (2013). Silent Films and Strange Stories: Theory of Mind, Gender, and Social Experiences in Middle Childhood. *Child Development*, 84(3), 989–1003. <http://doi.org/10.1111/cdev.12017>
- Devine, R. T., & Hughes, C. (2014). Relations Between False Belief Understanding and Executive Function in Early Childhood: A Meta-Analysis. *Child Development*, 85(5), 1777–1794. <http://doi.org/10.1111/cdev.12237>
- Dews, S., Winner, E., Kaplan, J., Rosenblatt, E., Hunt, M., Lim, K., ... Smarsh, B. (1996). Children's understanding of the meaning and functions of verbal irony. *Child Development*, 67(6), 3071–3085. <http://doi.org/10.2307/1131767>
- Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: cognitive functions, anatomy, and biochemistry. In *Principles of frontal lobe function* (D.T. Stuss, pp. 446–503). New York: Oxford University Press.
- Dumontheil, I. (2015). Development of the social brain during adolescence. *Psicología Educativa*, 21(2), 117–124. <http://doi.org/10.1016/j.pse.2015.08.001>

- Dumontheil, I., Apperly, I. a., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, 13, 331–338. <http://doi.org/10.1111/j.1467-7687.2009.00888.x>
- Dumontheil, I., Hillebrandt, H., Apperly, I. A., & Blakemore, S.-J. (2012). Developmental Differences in the Control of Action Selection by Social Information. *Journal of Cognitive Neuroscience*, 24, 2080–2095.
- Dumontheil, I., Küster, O., Apperly, I. a., & Blakemore, S. J. (2010). Taking perspective into account in a communicative task. *NeuroImage*, 52(4), 1574–1583. <http://doi.org/10.1016/j.neuroimage.2010.05.056>
- Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, 40(6), 760–768.
- Falk, E. ., Cascio, C. ., O'Donnell, M. ., Carp, J., Tinney, F. ., Bingham, C. R., ... Simons-Morton, B. . (2014). Neural responses to exclusion predict susceptibility to social influence. *Journal of Adolescent Health*, 54, S22–31. <http://doi.org/http://dx.doi.org/10.1016/j.jadohealth.2013.12.035>
- Ferguson, H. J., Apperly, I., Ahmad, J., Bindemann, M., & Cane, J. (2015). Task constraints distinguish perspective inferences from perspective use during discourse interpretation in a false belief task, 139, 50–70.
- Ferguson, H. J., & Breheny, R. (2011). Eye movements reveal the time-course of anticipating behaviour based on complex, conflicting desires. *Cognition*, 119(2), 179–196. <http://doi.org/10.1016/j.cognition.2011.01.005>
- Ferguson, H. J., & Breheny, R. (2012). Listeners' eyes reveal spontaneous sensitivity to others' perspectives. *Journal of Experimental Social Psychology*, 48(1), 257–263. <http://doi.org/10.1016/j.jesp.2011.08.007>
- Ferguson, H. J., Cane, J. E., Douchkov, M., & Wright, D. (2015). Empathy predicts false belief reasoning ability: evidence from the N400. *Social Cognitive and Affective Neuroscience*, 10(6), 848–855. <http://doi.org/10.1093/scan/nsu131>
- Fett, A. K. J., Shergill, S. S., Gromann, P. M., Dumontheil, I., Blakemore, S. J., Yakub, F., & Krabbendam, L. (2014). Trust and social reciprocity in adolescence - A matter of perspective-taking. *Journal of Adolescence*, 37(2), 175–184. <http://doi.org/10.1016/j.adolescence.2013.11.011>
- Filippova, E., Astington, J. W., Filippova, E., & Astington, J. W. (2008). Further Development in Social Reasoning Revealed in Discourse Irony Understanding. *Child Development*, 79, 126–138.
- Frith, C. D., & Frith, U. (2007). Social Cognition in Humans. *Current Biology*, 17(16), 724–732. <http://doi.org/10.1016/j.cub.2007.05.068>
- Frith, U. (2012). Why we need cognitive explanations of autism. *Quarterly Journal of Experimental Psychology*, 65(11), 2073–2092. <http://doi.org/10.1080/17470218.2012.697178>
- Frith, U., & de Vignemont, F. (2005). Egocentrism, allocentrism, and Asperger syndrome. *Consciousness and Cognition*, 14(4), 719–738. <http://doi.org/10.1016/j.concog.2005.04.006>

- Fuhrmann, D., Knoll, L. J., & Blakemore, S. (2015). Adolescence as a Sensitive Period of Brain Development. *Trends in Cognitive Sciences*, 19(10), 558–566. <http://doi.org/10.1016/j.tics.2015.07.008>
- Fuhrmann, D., Knoll, L. J., Sakhardande, A., Speekenbrink, M., Cohen Kadosh, K., & Blakemore, S. J. (2016). Perception and recognition of faces in adolescence. *Scientific Reports*, 6(33497), 1–8.
- Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of ???theory of mind??? *Trends in Cognitive Sciences*, 7(2), 77–83. [http://doi.org/10.1016/S1364-6613\(02\)00025-6](http://doi.org/10.1016/S1364-6613(02)00025-6)
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, 40, 177–190.
- Geangu, E., Gibson, A., Kaduk, K., & Reid, V. M. (2013). The neural correlates of passively viewed sequences of true and false beliefs. *Social Cognitive and Affective Neuroscience*, 8, 432–437.
- Gibbs, R. W. (1994). Figurative thought and figurative language. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 411–446). San Diego, CA: Academic Press.
- Gibbs, R. W. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics*, 34, 457–486.
- Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., ... Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neuroscience*, 2(10), 861–863. <http://doi.org/10.1038/13158>
- Giedd, J. N., & Rapoport, J. L. (2010). Structural MRI of Pediatric Brain Development: What Have We Learned and Where Are We Going? *Neuron*, 67(5), 728–734. <http://doi.org/10.1016/j.neuron.2010.08.040>
- Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, 7, 183–206.
- Goddings, A., Heyes, S. B., Bird, G., Viner, R. M., & Blakemore, S. (2012). The relationship between puberty and social emotion processing. *Developmental Science*, 15, 801–811. <http://doi.org/10.1111/j.1467-7687.2012.01174.x>
- Goldman, A. I., & Sebanz, N. (2005). Simulation, mirroring, and a different argument from error. *Trends in Cognitive Sciences*, 9(7), 320. <http://doi.org/10.1016/j.tics.2005.05.008>
- Graham, S., Bellmore, A., & Juvonen, J. (2003). Peer victimization in middle school: When self and peer views diverge. *Journal of Applied School Psychology*, 19, 117–137.
- Grice, P. H. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech Acts: Syntax and Semantics* (pp. 41–58). New York: Academic Press.
- Grossmann T, Johnson MH, Lloyd-Fox S, Blasi A, Deligianni F, et al. (2008) Early cortical specialization for face-to-face communication in human infants. *Proceedings of The Royal Society* 275: 2803–2811
- Grice, P. H. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, 20, 667–696.

- Gunther Moor, B., Op de Macks, Z. A., Güroglu, B., Rombouts, S. A. R. B., Van der Molen, M. W., & Crone, E. A. (2011). Neurodevelopmental changes of reading the mind in the eyes. *Social Cognitive and Affective Neuroscience*, 7, 44–52.
- Hadwin, J., & Perner, J. (1991). Pleased and surprised: Children's cognitive theory of emotion. *British Journal of Developmental Psychology*, 9, 215–234.
- Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, 18(10), 1279–1296.
- Hancock, J. T., Dunham, P. J., Purdy, K., Hancock, J. T., Dunham, P. J., & Purdy, K. (2000). Children's Comprehension of Critical and Complimentary Forms of Verbal Irony. *Journal of Cognition and Development*, 1(2), 227–248.  
<http://doi.org/10.1207/S15327647JCD010204>
- Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, 49(1), 43–61. [http://doi.org/10.1016/S0749-596X\(03\)00022-6](http://doi.org/10.1016/S0749-596X(03)00022-6)
- Happe, G E, F. (1993). Communicative competence and theory of mind in autism : A test of relevance theory. *Cognition*, 48, 101–119.
- Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able, mentally handicapped and normal children. *Journal of Autism and Developmental Disorders*, 24, 129–154.
- Harris, M., & Pexman, P. M. (2003). Children's perceptions of the social functions of verbal irony. *Discourse Processes*, 36, 147–165.
- Heller, A. S., & Casey, B. J. (2016). The neurodynamics of emotion: delineating typical and atypical emotional processes during adolescence. *Developmental Science*, 19(1), 3–18.  
<http://doi.org/10.1111/desc.12373>
- Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, 108(3), 831–836.  
<http://doi.org/10.1016/j.cognition.2008.04.008>
- Huttenlocher, P. R. (1979). Synaptic density in human frontal cortex - Developmental changes and effects of aging. *Brain Research*, 163(2), 195–205.
- Isaacs, J., Card, N. ., & Hodges, E. V. . (2001). Victimization by peers in the school context. *NYS Psychologist*, 13, 21–24.
- Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: implications for the acquisition of informativeness and implicature. *Cognition*, 120, 67–81.  
<http://doi.org/10.1016/j.cognition.2011.02.015>
- Katz, A., Blasko, D., & Kazamerski, V. (2004). Saying what you don't mean social influences on sarcastic language processing. *Current Directions in Psychological Science*, 13, 186–189.
- Katz, A. N. (2005). Discourse and sociocultural factors in understanding nonliteral language. In H. Colston & A. N. Katz (Eds.), *Figurative language comprehension: Social and cultural influences*. (pp. 183–207). Mahwah, NJ: Erlbaum.
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62(6), 593–602.

- Keulers, E. H. H., Evers, E. A. T., Stiers, P., & Jolles, J. (2010). Age, sex, and pubertal phase influence mentalizing about emotions and actions in adolescents. *Developmental Neuropsychology*, 35, 555–569.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Research Article TAKING PERSPECTIVE IN CONVERSATION : The Role of Mutual Knowledge in Comprehension, 11(1), 32–38.
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89(1), 25–41. [http://doi.org/10.1016/S0010-0277\(03\)00064-7](http://doi.org/10.1016/S0010-0277(03)00064-7)
- Kidd, D. C., & Castano, E. (2013). Reading Literary Fiction Improves Theory of Mind. *SCIENCE*, 342(October), 377–381.
- Kidd, D., Ongis, M., & Castano, E. (2016). On literary fiction and its effects on theory of mind. *Scientific Study of Literature*, 6, 42–58.
- Kilford, E. J., Garrett, E., & Blakemore, S. (2016). The Development of Social Cognition in Adolescence : An Integrated Perspective. *Neuroscience and Biobehavioral Reviews*, 70, 106–120. <http://doi.org/10.1016/j.neubiorev.2016.08.016>
- Kronmüller, E., & Barr, D. J. (2015). Referential precedents in spoken language comprehension: A review and meta-analysis. *Journal of Memory and Language*, 83, 1–19. <http://doi.org/10.1016/j.jml.2015.03.008>
- Lagattuta, H. K., Sayfan, L., & Harvey, C. (2014). Beliefs About Thought Probability: Evidence for Persistent Errors in Mindreading and Links to Executive Control. *Child Development*, 85(2), 659–674. <http://doi.org/10.1111/cdev.12154>
- Lagattuta, K. H., Sayfan, L., & Blattman, A. J. (2010). Forgetting common ground: six- to seven-year-olds have an overinterpretive theory of mind. *Developmental Psychology*, 46(6), 1417–32. <http://doi.org/10.1037/a0021062>
- Lebel, C., & Beaulieu, C. (2011). Longitudinal Development of Human Brain Wiring Continues from Childhood into Adulthood. *The Journal of Neuroscience*, 31(30), 10937–10947. <http://doi.org/10.1523/JNEUROSCI.5302-10.2011>
- Lecce, S., Bianco, F., Devine, R. T., Hughes, C., & Banerjee, R. (2014). Promoting theory of mind during middle childhood: a training program. *Journal of Experimental Child Psychology*, 126, 52–67. <http://doi.org/10.1016/j.jecp.2014.03.002>
- Lecce, S., Caputi, M., & Pagnin, A. (2014). Long-term effect of theory of mind on academic achievement: The role of sensitivity to criticism. *European Journal of Developmental Psychology*, 11, 305–318.
- Leon-Carrion, J., García-Orza, J., & Pérez-Santamaría, F. J. (2004). Development of the inhibitory component of the executive functions in children and adolescents. *The International Journal of Neuroscience*, 114(10), 1291–311. <http://doi.org/10.1080/00207450490476066>
- Lev-Wiesel, R., Nuttman-Shwartz, O., & Sternberg, R. (2006). Peer rejection during adolescence: Psychological long-term effects – a brief report. *Journal of Loss & Trauma*, 11, 131–142.
- Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, 46(3), 551–556. <http://doi.org/10.1016/j.jesp.2009.12.019>
- Luciana, M., Conklin, H. M., Hooper, C. J., & Yarger, R. S. (2005). The development of nonverbal

- working memory and executive control processes in adolescents. *Child Development*, 76, 697–712.
- Luck, S. L. (2014). *An Introduction to the Event-Related Potential Technique* (Second Edi). Cambridge, Massachusetts: MIT Press.
- Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004a). Maturation of Cognitive Processes From Late Childhood to Adulthood, 75(5), 1357–1372.
- Luna, B., Garver, K. E., Urban, T. a, Lazar, N. a, & Sweeney, J. a. (2004b). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, 75(5), 1357–72. <http://doi.org/10.1111/j.1467-8624.2004.00745.x>
- Luna, B., Padmanabhan, A., & Hearn, K. O. (2011). What has fMRI told us about the Development of Cognitive Control, 72(1), 1–28. <http://doi.org/10.1016/j.bandc.2009.08.005.What>
- Luria, A. R. (1966). *Higher cortical functions in man*. New York: Basic Books.
- McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning reexplored. *Cognition*, 102, 179–218.
- Meinhardt, J., Kühn-Popp, N., Sommer, M., & Sodian, B. (2012). Distinct neural correlates underlying pretense and false belief reasoning: evidence from ERPs. *NeuroImage*, 63, 623–631.
- Mills, K. L., Clasen, L. S., Giedd, J. N., & Blakemore, S. (2014). Developmental changes in the structure of the social brain in late childhood and adolescence. *Social Cognitive and Affective Neuroscience*, 9, 123–131. <http://doi.org/10.1093/scan/nss113>
- Monk, C. S., McClure, E. B., Nelson, E. E., Zarah, E., Bilder, R., Leibenluft, E., ... Pine, D. S. (2003). Adolescent immaturity in attention-related brain engagement to emotional facial expressions. *NeuroImage*, 20, 420–428.
- Nadig, A. S., & Sedivy, J. C. (2002). Evidence of Perspective-Taking Constraints in Children's On-Line Reference Resolution. *Psychological Science*, 13(4), 329–336. <http://doi.org/10.1111/1467-9280.00460>
- Nakassis, C., & Snedeker, J. (2002). Beyond sarcasm: Intonation and context as relational cues in children's recognition of irony. In *Proceedings of the 26th Boston University conference on language development*. (A. Greenhi). Somerville, MA: Cascadilla Press.
- Nicholson, A., Whalen, J. M., & Pexman, P. M. (2013). Children's processing of emotion in ironic language, 4(October), 1–10. <http://doi.org/10.3389/fpsyg.2013.00691>
- Nieuwland, M. S., & Van Berkum, J. J. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, 24(3), 691–701. <http://doi.org/10.1016/j.cogbrainres.2005.04.003>
- Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, 58(2), 220–249.
- Noveck, I. A. (2001). When Children are More Logical Than Adults: Experimental Investigations of Scalar Implicature. *Cognition*, 78, 165–188.
- O'Brien, S. F., & Bierman, K. L. (1988). Conceptions and perceived influence of peer groups - interviews with preadolescents and adolescents. *Child Development*, 59, 1360–1365.

- Onishi, K., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308, 255–258.
- Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition*, 86, 253–282.
- Paus, T., Keshavan, M., & Giedd, J. . (2008). Why do many psychiatric disorders emerge during adolescence? *Nature Reviews Neuroscience*, 9(12), 947–957.  
<http://doi.org/doi:10.1038/nrn2513>
- Peake, S. J., Dishion, T. J., Stormshak, E. A., Moore, W. E., & Pfeifer, J. H. (2013). Risk-taking and social exclusion in adolescence: neural mechanisms underlying peerinfluences on decision-making. *NeuroImage*, 82, 23–34.  
<http://doi.org/http://dx.doi.org/10.1016/j.neuroimage.2013.05.061>
- Pearson, B. Z. (1990). The comprehension of metaphor by preschool children. *Journal of Child Language*, 17, 185–203.
- Petanjek, Z., Judaš, M., Šimic, G., Rasin, M. R., Uylings, H. B. M., Rakic, P., & Kostovic, I. (2011). Extraordinary neoteny of synaptic spines in the human prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 108(32), 13281–13286.
- Pexman, P. M. (2008). It's fascinating research: the cognition of verbal irony. *Current Directions in Psychological Science*, 17(4), 286–290.
- Pexman, P. M., & Glenwright, M. (2007). How do typically developing children grasp the meaning of verbal irony ? *Journal of Neurolinguistics*, 20, 178–196.  
<http://doi.org/10.1016/j.jneuroling.2006.06.001>
- Pexman, P. M., Glenwright, M., Krol, A., James, T., Pexman, P. M., Glenwright, M., ... James, T. (2005). An Acquired Taste : Children's Perceptions of Humor and Teasing in Verbal Irony. *Discour*, 40(3), 259–288. <http://doi.org/10.1207/s15326950dp4003>
- Pfeifer, J. H., Lieberman, M. D., & Dapretto, M. (2007). I know you are but what am I!": Neural bases of self and social knowledge retrieval in children and adults. *Journal of Cognitive Neuroscience*, 19, 1323–1337.
- Pfeifer, J. H., Masten, C. L., Borofsky, L. A., Dapretto, M., & Lieberman, M. D. (2011). Neural correlates of direct and reflected self-appraisals in adolescents and adults: when social perspective-taking informs self-perception. *Child Development*, 80(4), 1016–1038.  
<http://doi.org/10.1111/j.1467-8624.2009.01314.x>Neural
- Pfeifer, J. H., Masten, C. L., Moore, W. E., Oswald, T. M., Mazziotta, J. C., Iacoboni, M., & Dapretto, M. (2011). Entering adolescence: resistance to peer influence, risky behavior, and neural changes in emotion reactivity. *Neuron*, 69(5), 1029–1036.  
<http://doi.org/10.1016/j.neuron.2011.02.019>
- Pouscoulous, N., Noveck, I. a., Politzer, G., & Bastide, a. (2007). A developmental investigation of processing costs in implicature production, (February 2015), 37–41.  
<http://doi.org/10.1080/10489220701600457>
- Prinstein, M. ., Sheah, C. ., & Guyer, A. . (2005). Peer victimization, cue interpretation, and internalizing symptoms: Preliminary concurrent and longitudinal findings for children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, 34, 11–24.
- R Core Development Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.r->



- Regel, S., Coulson, S., & Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain Research, 1311*, 121–135. <http://doi.org/10.1016/j.brainres.2009.10.077>
- Regel, S., & Gunter, T. C. (2016). What exactly do you mean? ERP evidence on the impact of explicit cueing on language comprehension. In *Pre-proceedings of Trends in Experimental Pragmatics* (pp. 115–120).
- Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn't It Ironic ? An Electrophysiological Exploration of Figurative Language Processing. *Journal of Cognitive Neuroscience, 23*(2), 277–293. <http://doi.org/10.1162/jocn.2010.21411>
- Regel, S., Meyer, L., & Gunter, T. C. (2014). Distinguishing neurocognitive processes reflected by P600 effects: evidence from ERPs and neural oscillations. *PloS One, 9*(5), e96840. <http://doi.org/10.1371/journal.pone.0096840>
- Rigby, K. (2000). Effects of peer victimization in schools and perceived social support on adolescent well-being. *Journal of Adolescence, 23*, 57–60.
- Rigby, K. (2003). Consequences of bullying in schools. *Canadian Journal of Psychiatry, 48*, 583–590.
- Rubio-Fernández, P. (2016). The director task: A test of Theory-of-Mind use or selective attention? *Psychonomic Bulletin & Review*. <http://doi.org/10.3758/s13423-016-1190-7>
- Rubio-Fernández, P., & Glucksberg, S. (2012). Reasoning about other people's beliefs: Bilinguals have an advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(1), 211–217. <http://doi.org/10.1037/a0025162>
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology. Human Perception and Performance, 36*, 1255–1266. <http://doi.org/10.1037/a0018729>
- Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology, 16*(2), 235–239. <http://doi.org/10.1016/j.conb.2006.03.001>
- Schneider, D., Bayliss, A. P., Becker, S., & Dux, P. E. (2012). Sustained implicit belief processing revealed by eye movements. *Experimental Psychology: General, 141*, 433–438.
- Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts theory of mind processing. *Psychological Science, 23*, 842–847.
- Schneider, D., & Low, J. (2016). Efficient versus flexible mentalizing in complex social settings: Exploring signature limits. *British Journal of Psychology, 107*, 26–29.
- Schneider, D., Nott, Z. E., & Dux, P. E. (2014). Task instructions and implicit theory of mind. *Cognition, 133*(1), 43–47. <http://doi.org/10.1016/j.cognition.2014.05.016>
- Schneider, D., Slaughter, V. P., & Dux, P. E. (2015). What do we know about implicit false-belief tracking? *Psychonomic Bulletin & Review, 22*, 1–12.
- Schneider, D., Slaughter, V. P., & Dux, P. E. (2017). Current evidence for automatic Theory of Mind processing in adults. *Cognition, 162*, 27–31.
- Sebastian, C. L., Fontaine, N. M. G., Bird, G., Blakemore, S.-J., Brito, S. a De, McCrory, E. J. P., &

- Viding, E. (2012). Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, 7(1), 53–63. <http://doi.org/10.1093/scan/nsr023>
- Sebastian, C. L., Tan, G. C. Y., Roiser, J. P., Viding, E., Dumontheil, I., & Blakemore, S. (2011). Developmental influences on the neural bases of responses to social rejection: Implications of social neuroscience for education. *NeuroImage*, 57, 686–694. <http://doi.org/10.1016/j.neuroimage.2010.09.063>
- Sebastian, C. L., Viding, E., Williams, K. D., & Blakemore, S. (2010). Social brain development and the affective consequences of ostracism in adolescence. *Brain and Cognition*, 72, 134–145. <http://doi.org/10.1016/j.bandc.2009.06.008>
- Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, 298, 199–209. <http://doi.org/doi:10.1098/rstb.1982.0082>
- Shamay-Tsoory, S. G., Harari, H., Aharon-Peretz, J., & Levkovitz, Y. (2010). The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex*, 46(5), 668–677. <http://doi.org/10.1016/j.cortex.2009.04.008>
- Shamay-Tsoory, S. G., Harari, H., Aharon-Peretz, J., & Levkovitz, Y. (2010). The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex*, 46, 668–677.
- Shamay-Tsoory, S. G., Tomer, R., & Aharon-Peretz, J. (2005). The Neuroanatomical Basis of Understanding Sarcasm and Its Relationship to Social Cognition. *Neuropsychology*, 19(3), 288–300. <http://doi.org/10.1037/0894-4105.19.3.288>
- Shaw, P., Kabani, N. ., Lerch, J. ., Eckstrand, K., Lenroot, R., Gogtay, N., ... Wise, S. . (2008). Neurodevelopmental trajectories of the human cerebral cortex. *Journal of Neuroscience*, 28, 3586–3594.
- Shibata, M., Terasawa, Y., Osumi, T., Masui, K., Ito, Y., Sato, A., & Umeda, S. (2017). Time course and localization of brain activity in humor comprehension: an ERP/sLORETA study. *Brain Research*, 1657, 215–222. <http://doi.org/10.1016/j.brainres.2016.12.010>
- Shibata, M., Toyomura, A., Itoh, H., & Abe, J. (2010). Neural substrates of irony comprehension: a functional MRI study. *Brain Research*, 1308, 114–123.
- Simmonds, D. J., Pekar, J. J., & Mostofsky, S. H. (2008). Meta-analysis of Go/No-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia*, 46(1), 224–32. <http://doi.org/10.1016/j.neuropsychologia.2007.07.015>
- Singer, T., Critchley, H. ., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences*, 13, 334–340.
- Slaughter, V., Imuta, K., Peterson, C. C., & Henry, J. D. (2015). Meta-analysis of theory of mind and peer popularity in the preschool and early school years. *Child Development*, 86, 1159–1184. <http://doi.org/http://dx.doi.org/10.1111/cdev.12372>
- Sommerville, J. A., Bernstein, D. M., & Meltzoff, A. N. (2013). Measuring false belief in centimeters: Adults and children fail to suppress privileged knowledge on a novel change-of-location task. *Child Development*, 84, 1846–1854.

- Sowell, E. ., Thompson, P. ., Holmes, C. ., Jernigan, T. ., & Toga, A. . (1999). In vivo evidence for post-adolescent brain maturation in frontal and striatal regions. *Nature Neuroscience*, 2(10), 859–861.
- Spotorno, N., Cheylus, A., Van Der Henst, J.-B., & Noveck, I. a. (2013). What's behind a P600? Integration Operations during Irony Processing. *PLoS ONE*, 8(6), e66839. <http://doi.org/10.1371/journal.pone.0066839>
- Spotorno, N., Koun, E., Prado, J., Van Der Henst, J. B., & Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, 63(1), 25–39. <http://doi.org/10.1016/j.neuroimage.2012.06.046>
- Spotorno, N., & Noveck, I. a. (2014). When is irony effortful? *Journal of Experimental Psychology. General*, 143(4), 1649–1665. <http://doi.org/10.1037/a0036630>
- Sullivan, K., Winner, E., & Hopfield, N. (1995). How children tell a lie from a joke: The role of second-order mental state attributions. *British Journal of Developmental Psychology*, 13, 191–204.
- Surian, L., Surian, L., Caldi, S., Caldi, S., Sperber, D., & Sperber, D. (2007). Attribution of beliefs by 13- month-old infants. *Psychological Science*, 18(7), 580–586.
- Surtees, A. D. R., & Apperly, I. a. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development*, 83(2), 452–60. <http://doi.org/10.1111/j.1467-8624.2011.01730.x>
- Symeonidou, I., Dumontheil, I., Chow, W.-Y., & Breheny, R. (2016). Development of online use of theory of mind during adolescence: An eye-tracking study. *Journal of Experimental Child Psychology*, 149, 81–97. <http://doi.org/10.1016/j.jecp.2015.11.007>
- Tamnes, C. K., Walhovd, K. B., Dale, A. M., Østby, Y., Grydeland, H., Richardson, G., ... Fjell, A. M. (2013). Brain development and aging: Overlapping and unique patterns of change. *NeuroImage*, 68, 63–74. <http://doi.org/10.1016/j.neuroimage.2012.11.039>
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science*, 268(5217), 1632–1634.
- Thomas, L. A., De Bellis, M. D., Graham, R., & LaBar, K. (2007). Development of emotional face recognition in late childhood and adolescence. *Developmental Science*, 10(5), 547–558.
- Van den Brink D, Van Berkum JJ, Bastiaansen MCM, Tesink CMJY, Kos M, et al. (2010) Empathy matters: ERP evidence for inter-individual differences in social language processing. *Social cognitive and affective neuroscience* 7: 173– 183.
- van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22, 241–255.
- Vetter, N. C., Altgassen, M., Phillips, L., Mahy, C. E. V, & Kliegel, M. (2013). Development of affective theory of mind across adolescence: disentangling the role of executive functions. *Developmental Neuropsychology*, 38(2), 114–25. <http://doi.org/10.1080/87565641.2012.733786>
- Vetter, N. C., Leipold, K., Kliegel, M., Phillips, L. H., Vetter, N. C., Leipold, K., ... Altgassen, M. (2013). Ongoing development of social cognition in adolescence in adolescence. *Child Neuropsychology*, 19(6), 615–629. <http://doi.org/10.1080/09297049.2012.718324>
- Völlm, B. A., Taylor, A. N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., ... Elliot, R. (2006).

Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *NeuroImage*, 29(1), 90–98.

Waggoner, J. E., & Palermo, D. S. (1989). Betty is a bouncing bubble: Children's comprehension of emotion-descriptive metaphors. *Developmental Psychology*, 25, 152.

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Developmental changes in the neural basis of interpreting communicative intent. *Social Cognitive and Affective Neuroscience*, 1(2), 107–121. <http://doi.org/10.1093/scan/nsi018>

Wang, J. J., Ali, M., Frisson, S., & Apperly, I. a. (2015). Language complexity modulates 8- and 10-year-olds' success at using their theory of mind abilities in a communication task. *Journal of Experimental Child Psychology*. <http://doi.org/10.1016/j.jecp.2015.09.006>

Watanabe, J., Sugiura, M., Sato, K., Sato, Y., Maeda, Y., Matsue, Y., ... Kawashima, R. (2002). The Human Prefrontal and Parietal Association Cortices Are Involved in NO-GO Performances: An Event-Related fMRI Study. *NeuroImage*, 17(3), 1207–1216. <http://doi.org/10.1006/nimg.2002.1198>

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale – Third Edition Manual* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: Harcourt Assessment.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, 72(3), 655–684. <http://doi.org/10.1111/1467-8624.00304>

Westley, A., Kohút, Z., & Rueschemeyer, S. (2017). Journal of Experimental Child “‘ I know something you don ’ t know ’”: Discourse and social context effects on the N400 in adolescents. *Journal of Experimental Child Psychology*, 164, 45–54. <http://doi.org/10.1016/j.jecp.2017.06.016>

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128.

Winner, E., & Leekam, S. (1991). Distinguishing irony from deception: Understanding the speaker's second-order intention. *British Journal of Developmental Psychology*, 9, 257–270.

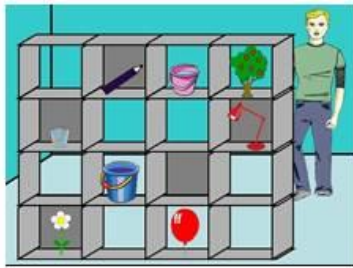
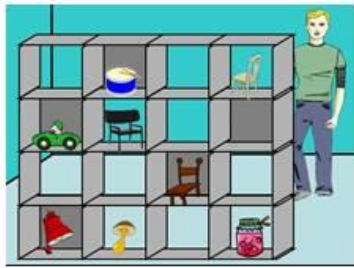
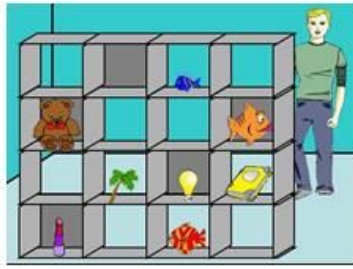
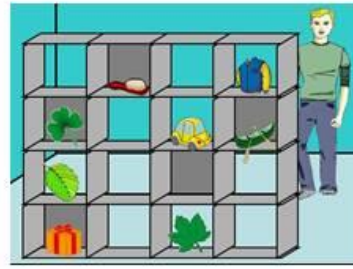
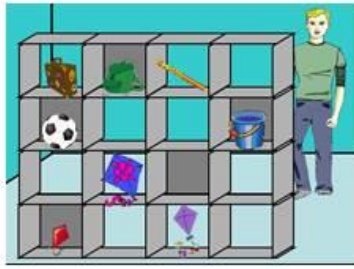
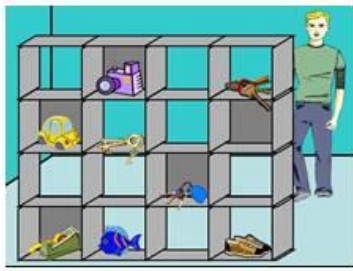
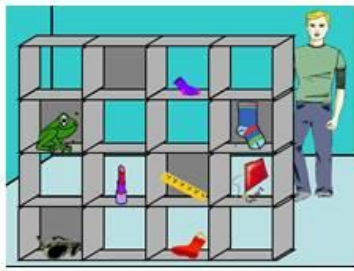
Winner, E., Windmueller, G., Rosenblatt, E., Bosco, L., Best, E., & Gardner, H. (1987). Making sense of literal and nonliteral falsehood. *Metaphor & Symbolic Activity*, 2, 13–32.

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science*, 18(7), 600–6. <http://doi.org/10.1111/j.1467-9280.2007.01946.x>

Yakovlev, P. A., & Lecours, I. R. (1967). The myelogenetic cycles of regional maturation of the brain. In *Regional Development of the brain in early life*. (A. Minkows). Oxford: Blackwell.







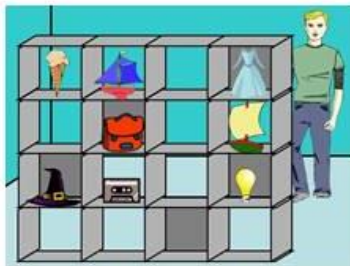
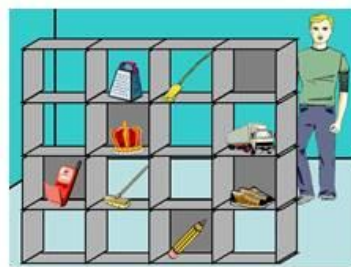
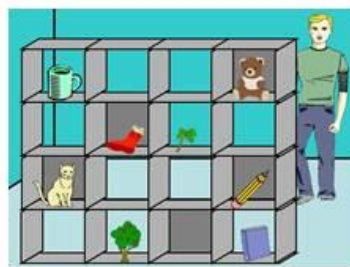
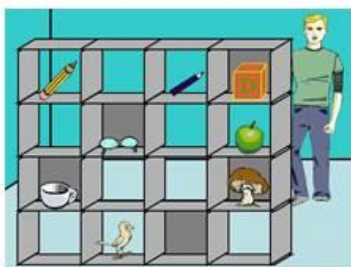
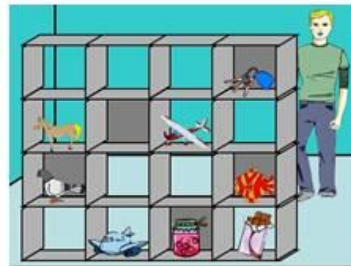
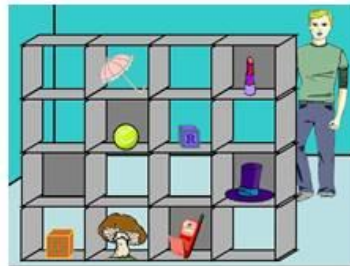
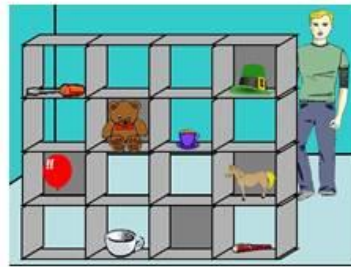
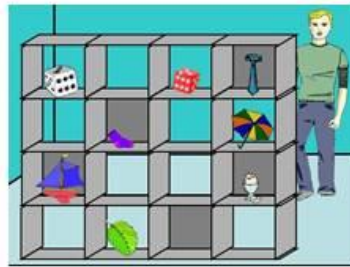
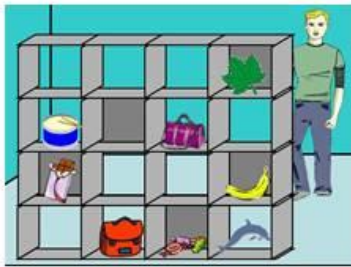
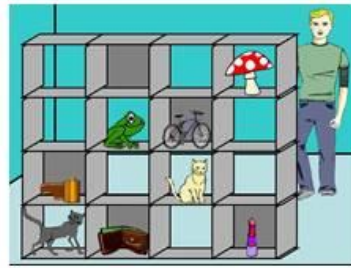
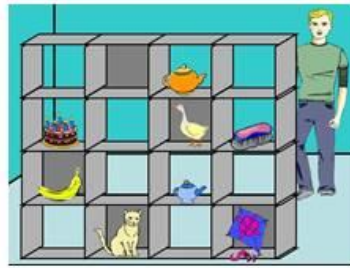
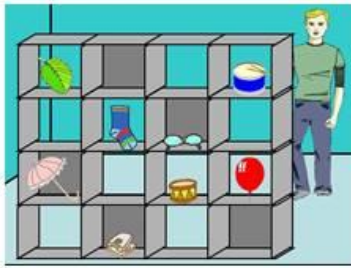
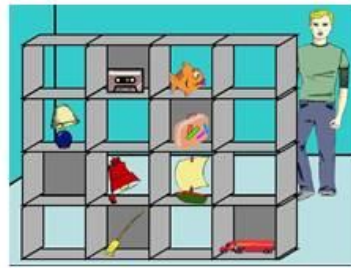
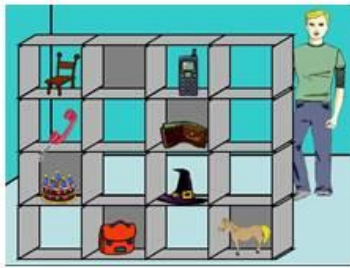
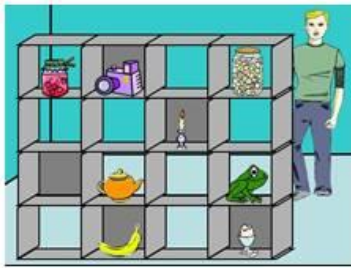










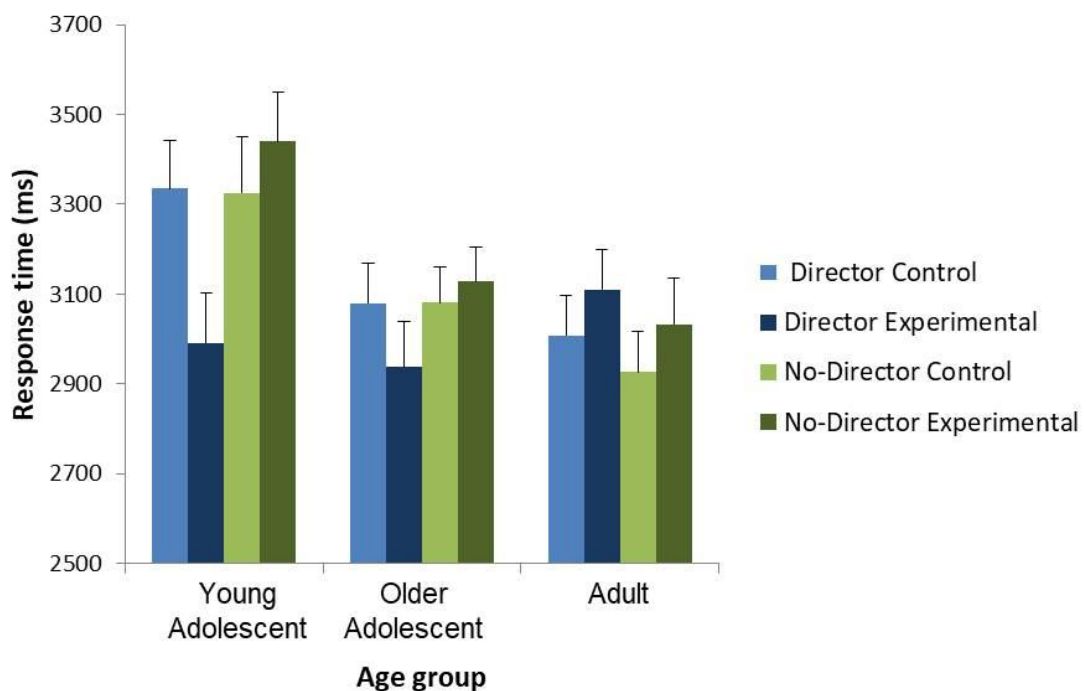


## Behavioural results

### Director task

#### Response time

A 2 x 2 x 3 mixed repeated measures ANOVA with Condition (Director, No-Director), Trial type (Control, Experimental) and Age group was performed on RT for all responses (**Figure S1**). The main effects of Condition ( $F(1, 57) = 2.49, p = .120$ ), Trial type ( $F(1, 57) = .343, p = .560$ ), and Age ( $F(2, 57) = 2.40, p = .100$ ) were not significant. There was however, a significant interaction between Trial type and Age ( $F(2, 57) = 3.66, p = .032, \eta_p^2 = .114$ ), and a trend for an interaction between Condition and Age ( $F(2, 57) = 2.78, p = .071, \eta_p^2 = .089$ ). Finally, there was a significant three-way interaction between Condition, Trial type and Age ( $F(2, 57) = 4.24, p = .019, \eta_p^2 = .130$ ), which was explored further by looking at Experimental and Control trials separately.



**Figure S1** Response time (mean + SE) in Control and Experimental trials of the Director and No-Director conditions for each age group.

A 2 x 3 ANOVA with Condition (Director, No-Director) and Age group showed a significant main effect of Age group ( $F(2, 57) = 4.09, p = .022, \eta_p^2 = .125$ ) in Control trials, but no significant interaction between Condition and Age ( $F(2, 57) = .228, p = .797$ ). Post-hoc comparisons showed younger adolescents responded slower to Control trials than



adults ( $p = .019$ ) but did not differ from older adolescents ( $p = .134$ ). Older adolescents and adults did not differ ( $p = .921$ ).

Conversely, in Experimental trials the main effect of Age was not significant ( $F(2, 57) = 1.09, p = .344$ ), but there was a significant interaction between Condition and Age ( $F(2, 57) = 5.92, p = .005, \eta_p^2 = .172$ ). This interaction was followed-up by considering Director and No-Director Experimental trials separately using one-way ANOVAs testing for the effect of Age on RT. In Director Experimental trials there was no significant effect of Age ( $F(2, 57) = 0.80, p = .453$ ), while there was a significant effect of Age in No-Director Experimental trials ( $F(2, 57) = 4.14, p = .021, \eta_p^2 = .123$ ). Post hoc t-tests showed that in the No-Director Experimental trials younger adolescents were slower than older adolescents ( $t(32) = 2.38, p = .023, d = .772$ ) and adults ( $t(32) = 2.66, p = .012, d = .939$ ), who did not differ from each other ( $t(44) = .77, p = .445$ ).

The analyses were rerun with RT for correct trials only (Young Adolescent,  $N=13$ , Older Adolescent,  $N=26$ ) and Adult group,  $N=20$ ). The interactions with age groups were similar, although the three-way interaction was at trend level only ( $F(2, 56) = 2.53, p = .089$ ).

### *Inhibitory control tasks*

#### *Response Time of Go trials*

One-way ANOVAs were performed on RT for correct Go trials to investigate the effect of Age in the Simple and Complex tasks separately. The main effect of Age was significant in Simple Go trials ( $F(2,57) = 12.64, p < .001, \eta_p^2 = .307$ ). Post-hoc comparisons showed that younger adolescents (mean = 434ms) were significantly slower than older adolescents (mean = 382ms;  $p < .001$ ) and adults (mean = 375ms;  $p < .001$ ), who did not differ ( $p = 1$ ). In Complex Go trials, the main effect of Age was significant ( $F(2,57) = 5.28, p = .008, \eta_p^2 = .156$ ). Again, post-hoc comparisons showed that younger adolescents (mean = 478ms) were slower than older adolescents ( $p = .022$ ) and adults ( $p = .011$ ). Older adolescents (mean = 423ms) and adults (mean = 415ms) did not differ ( $p = 1$ ).

### **Eye-tracking Results**

As mentioned in the main paper, for some models of perspective-taking, such as Keysar and colleagues' Perspective-Adjustment model, an error represents a failure to adjust from an initial egocentric perspective to a speaker-oriented perspective. Thus

analysis of correct and incorrect trials should reveal in all cases an initial bias to the privileged competitor. However, for other models (Brown-Schmidt & Hanna, 2011; Hanna et al., 2003) an incorrect trial may reflect a simple failure to integrate perspective from the start. If this is the case, patterns of eye gaze may differ from the onset of the critical phase of the instruction between correct and incorrect trials. Since an analysis of all data may be confounded by the differences in rates of error between age-groups, we performed a set of analyses that separates correct and incorrect trials as the main analysis. However, we report the analysis of all trials here for transparency.

#### *All trials data*

A mixed 2 x 2 x 3 repeated measures ANOVA with Condition (Director, No-Director), Trial type (Control, Experimental) and Age group was performed on target advantage for each region (see **Table S1** for mean target advantage scores across conditions, **Table S2** for omnibus ANOVA and **Figure S2a-d** for average target advantage during a trial for each age group across conditions). There were no significant main effects of Age in any of the regions (all  $ps > .50$ ). Regions 3 (“large”), 4 (“ball”), and 5 (“up”) showed a significant main effect of Condition (all  $ps < .05$ ), whereby the target advantage was negative and lower in the No-Director condition compared to the Director condition. The main effect of Trial type was significant in Region 3 ( $p < .01$ ,  $\eta_p^2 = .080$ ) and Region 5 ( $p < .01$ ,  $\eta_p^2 = .134$ ) but not Region 4 (“ball”) ( $p = .291$ ). In Region 3 target advantage was greater for Experimental trials while the opposite pattern was observed in Region 5. None of the Regions showed significant interactions between Age group and Condition (all  $ps > .180$ ), while only Region 5 (“up”) showed a significant interaction between Age group and Trial type ( $p = .017$ ), such that adults had a greater target advantage in Control trials whereas younger and older adolescents had greater target advantage in Experimental trials. The interaction between Condition and Trial type was significant for all Regions (all  $ps < .01$ ). In Region 3 and 4 the effect of Trial type was bigger in the Director rather than in the No-Director condition but the opposite pattern was found in Region 5. The three-way interaction of Age, Trial type and Condition was not significant for Region 3 ( $p = .980$ ) and Region 5 ( $p = .166$ ).

**Supplementary Table S1** Mean (SE) target advantage scores in each condition for each region.

Condition & Trial type	Region 3 "large"	Region 4 "ball"	Region 5 "up"
Young Adolescent			
Director Control	-0.08 (.02)	-0.11 (.02)	-0.04 (.03)
Director Experimental	0.04 (.03)	0.05 (.04)	0.10 (.06)
No-Director Control	-0.05 (.02)	-0.07 (.02)	-0.04 (.03)
No-Director Experimental	0.01 (.03)	-0.14 (.04)	-0.25 (.04)
Older Adolescent			
Director Control	-0.03 (.02)	-0.10 (.02)	0.00 (.02)
Director Experimental	0.04 (.02)	0.01 (.03)	0.00 (.05)
No-Director Control	0.03 (.01)	-0.09 (.01)	0.01 (.02)
No-Director Experimental	-0.03 (.02)	-0.14 (.02)	-.23 (.03)
Adult			
Director Control	-0.04 (.01)	-0.10 (.02)	0.03 (.02)
Director Experimental	0.04 (.03)	-0.07 (.03)	-0.08 (.05)
No-Director Control	-0.03 (.02)	-0.07 (.02)	0.01 (.03)
No-Director Experimental	0.00 (.02)	-0.14 (.03)	-0.24 (.04)

**Supplementary Table S2** Omnibus repeated measures ANOVA F-values by subject of target advantage in each region.

	df	Region 3	Region 4	Region 5
Omnibus ANOVA				
Age group	2, 57	0.23	0.70	0.18
Condition	1, 57	4.14*	16.13**	25.51**
Trial type	1, 57	15.43**	1.14	46.77**
Age group x Condition	2, 57	0.65	1.75	1.01
Age group x Trial type	2, 57	1.28	1.56	4.39*
Condition x Trial type	1, 57	9.06**	48.88**	33.45**
Age group x Condition x Trial	2, 57	0.02	3.10^	1.86

\*\* $p < .01$

\* $p < .05$

^ $p < .1$

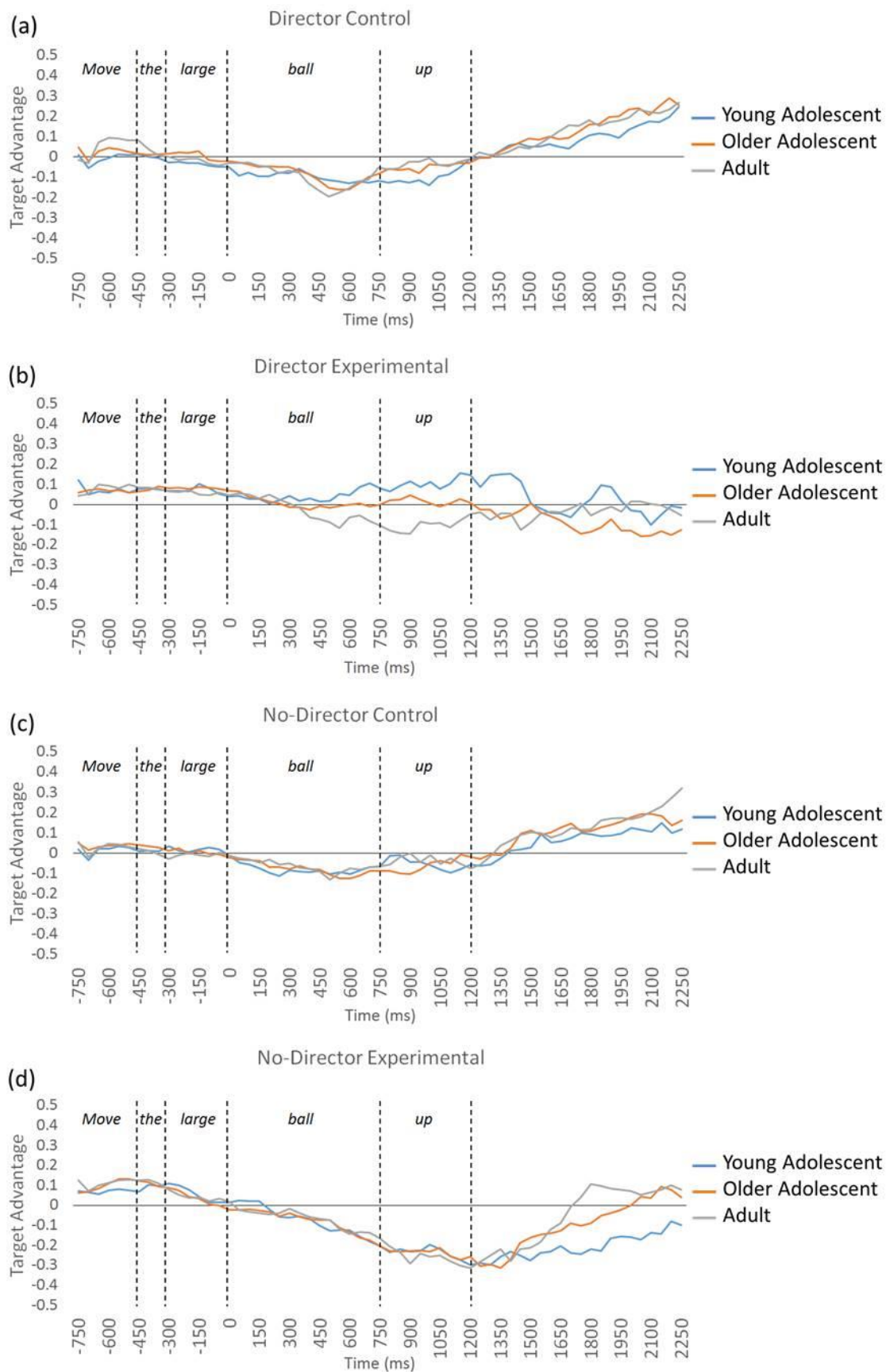
**Supplementary Table S3** Omnibus repeated measures ANOVA F-values by item of target advantage in each region.

	df	Region 3	Region 4	Region 5
Omnibus ANOVA				
Age group	2, 69	0.05	0.30	0.07
Condition	1, 69	2.98^	15.98**	27.36**
Trial type	1, 69	12.71**	1.19	32.52**
Age group x Condition	2, 69	0.26	1.79	1.33
Age group x Trial type	2, 69	1.23	2.13	3.92*
Condition x Trial type	1, 69	9.52**	45.45**	51.31**
Age group x Condition x Trial	2, 69	0.05	3.50*	2.67^

\*\* $p < .01$

\* $p < .05$

^ $p < .1$

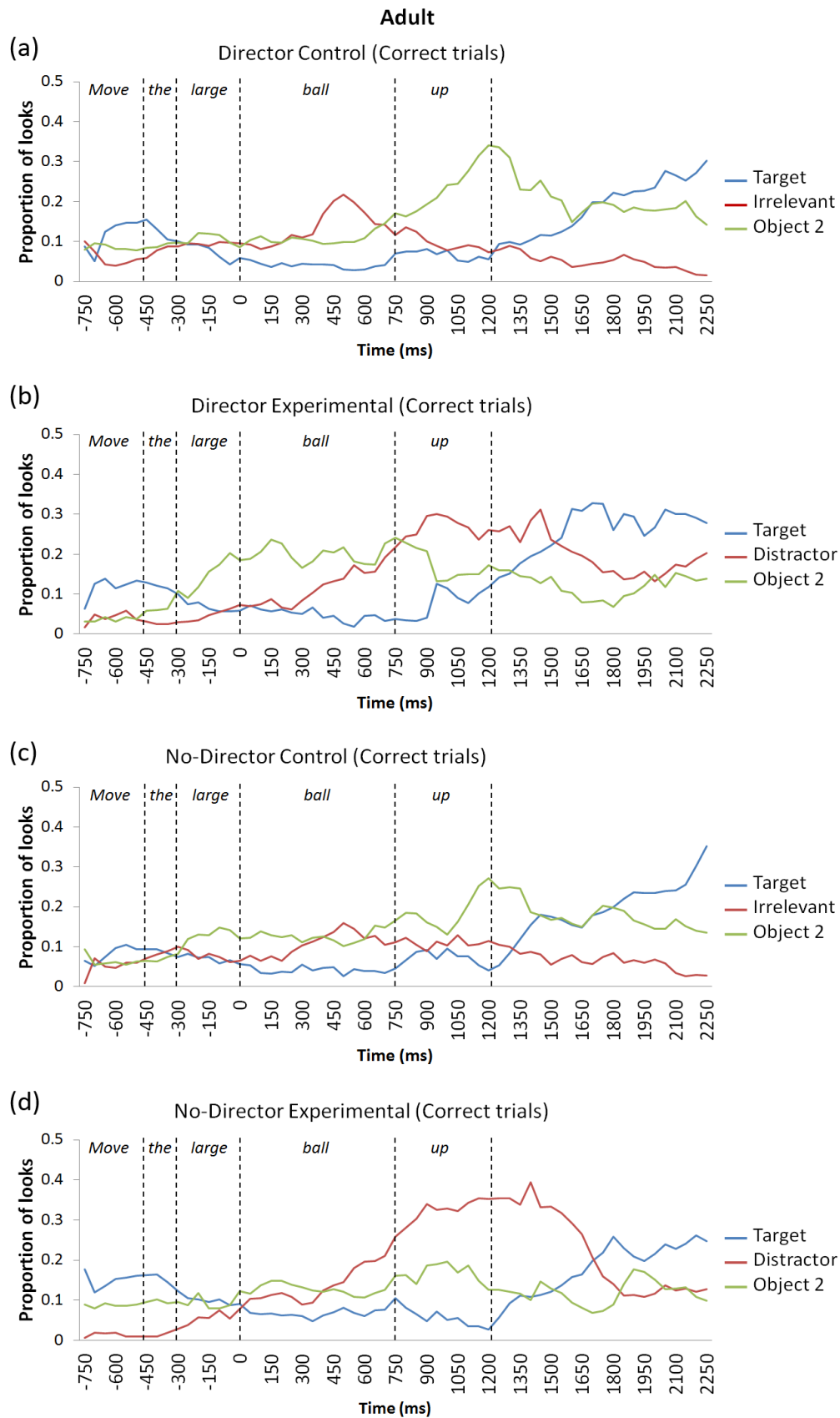


**Supplementary Figure S2a-d** Target advantage scores per condition over 50ms time bins.

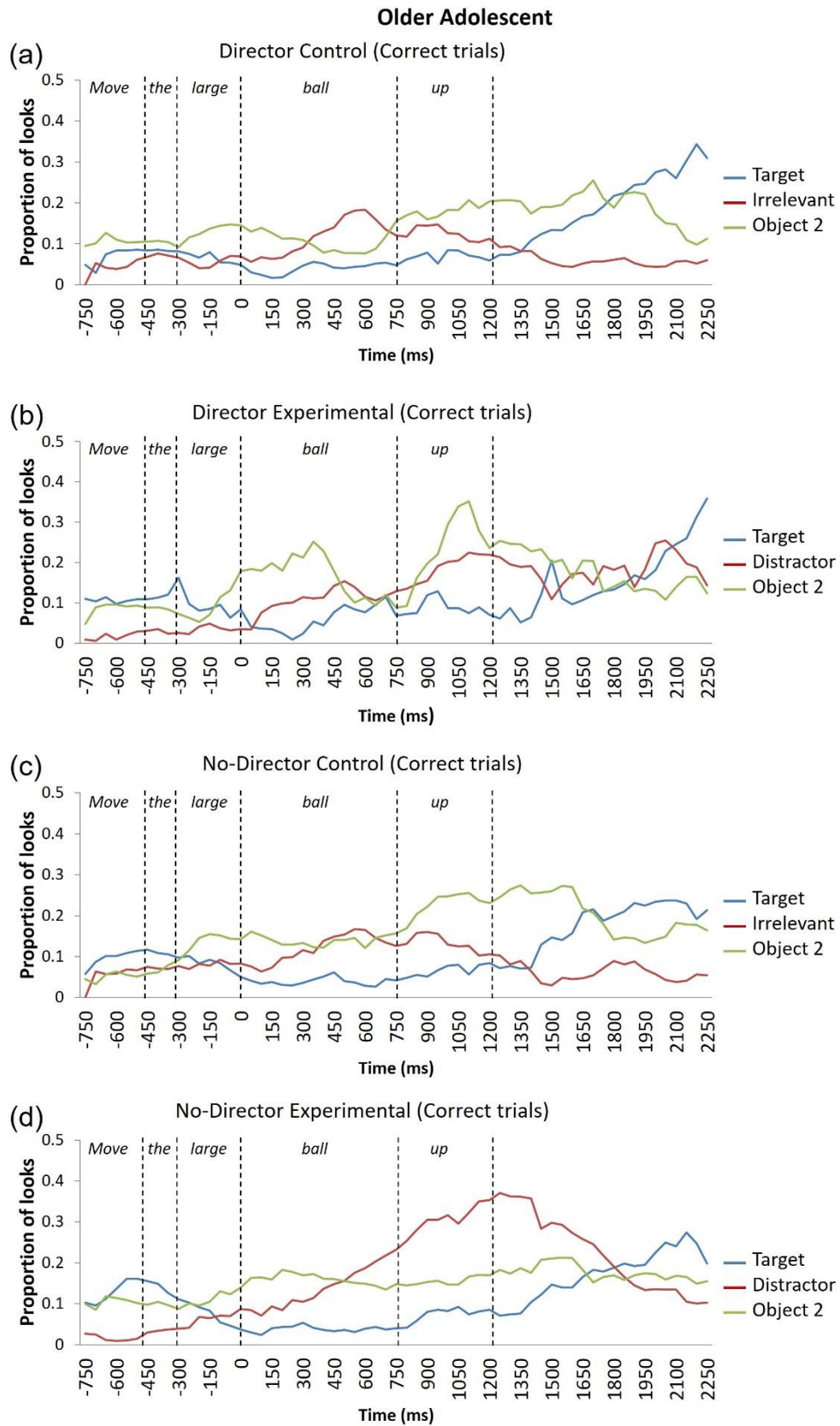


There was a marginally significant three-way interaction between Condition, Trial type and Age in Region 4 ( $p = .053$ ,  $\eta_p^2 = .016$ ). This interaction was explored further by analysing the Experimental and Control trials separately. A 2 x 3 ANOVA with Condition (Director, No-Director) and Age showed that for the Control trials neither the main effect of Age ( $F_1(2, 57) = .208$ ,  $p = .813$ ;  $F_2(2, 69) = .069$ ,  $p = .933$ ), nor the interaction between Condition and Age were significant ( $F_1(2,57) = 1.06$ ,  $p = .350$ ;  $F_2(2, 69) = .756$ ,  $p = .473$ ). In Experimental trials the main effect of Age was not significant ( $F_1(2, 57) = 1.34$ ,  $p = .27$ ;  $F_2(2, 69) = 1.20$ ,  $p = .306$ ), however there was a trend for an interaction between Condition and Age ( $F_1(2, 57) = 2.59$ ,  $p = .083$ ,  $\eta_p^2 = .034$ ;  $F_2(2, 69) = 3.33$ ,  $p = .041$ ,  $\eta_p^2 = .032$ ). The trend interaction was followed up with one-way ANOVA to explore the effect of age on Director Experimental trials and No-Director Experimental trials separately. Target advantage did not differ with Age in No-Director Experimental trials ( $F_1(2, 57) = .017$ ,  $p = .982$ ;  $F_2(2, 69) = .021$ ,  $p = .979$ ) but there was marginal effect of Age in the Director Experimental trials ( $F_1(2, 57) = 2.90$ ,  $p = .063$ ,  $\eta_p^2 = .092$ ;  $F_2(2, 69) = 6.62$ ,  $p = .002$ ,  $\eta_p^2 = .161$ ) where both adolescent groups had a bigger advantage score than adults.

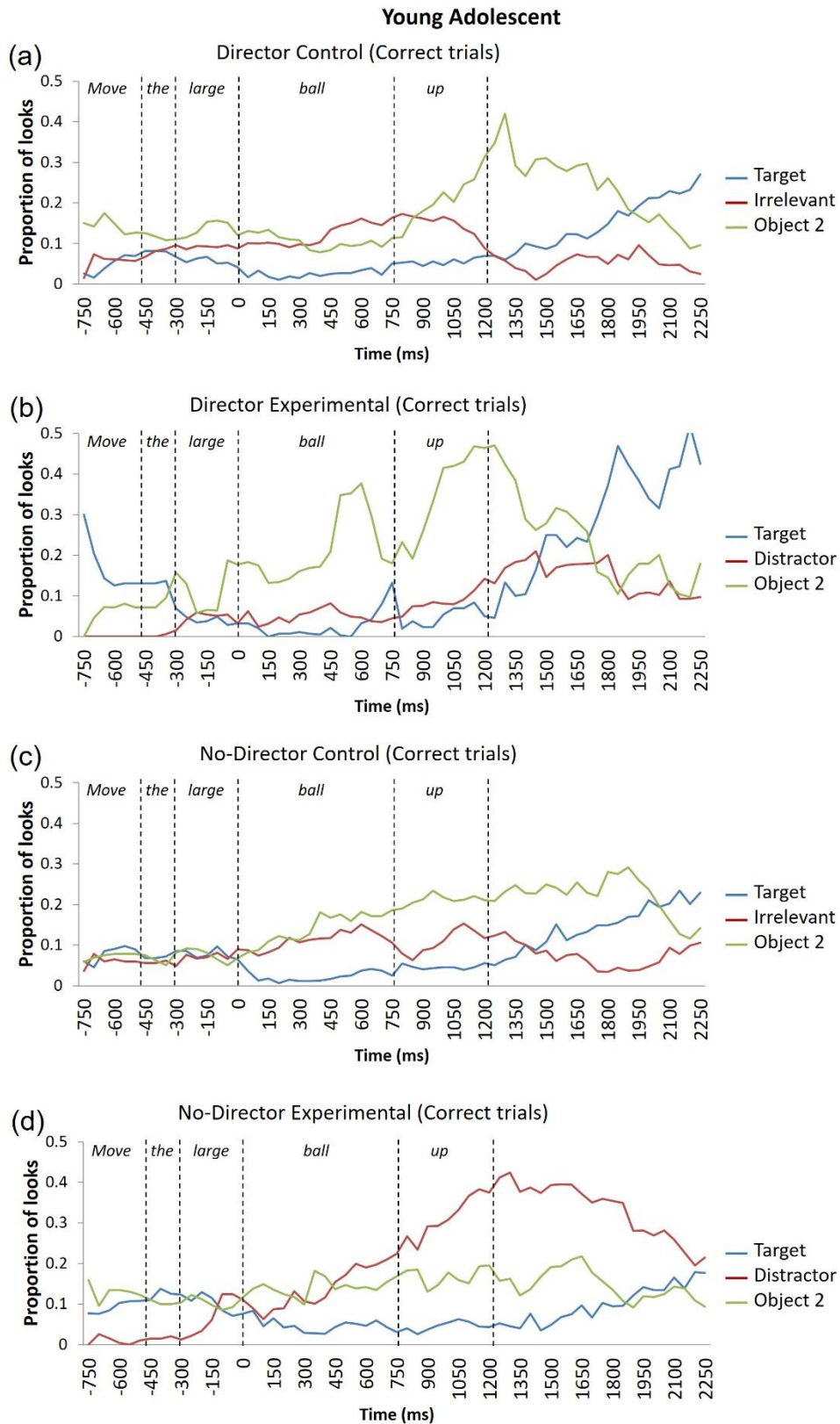
The marginal main effect of age in the Director Experimental trials in Region 4 showed that younger and older adolescents had a bigger target advantage than adults. This is contradictory to the behavioural results as younger and older adolescents made significantly more errors than the adults by choosing the distractor and therefore, one would predict a bigger target advantage in adults. As accuracy differed between groups, further eye-tracking analyses separated correct and incorrect trials (see main text).



**Supplementary Figure S3a-d** Proportion of looks to the target, distractor/irrelevant object and object 2 per condition over 50ms time bins.

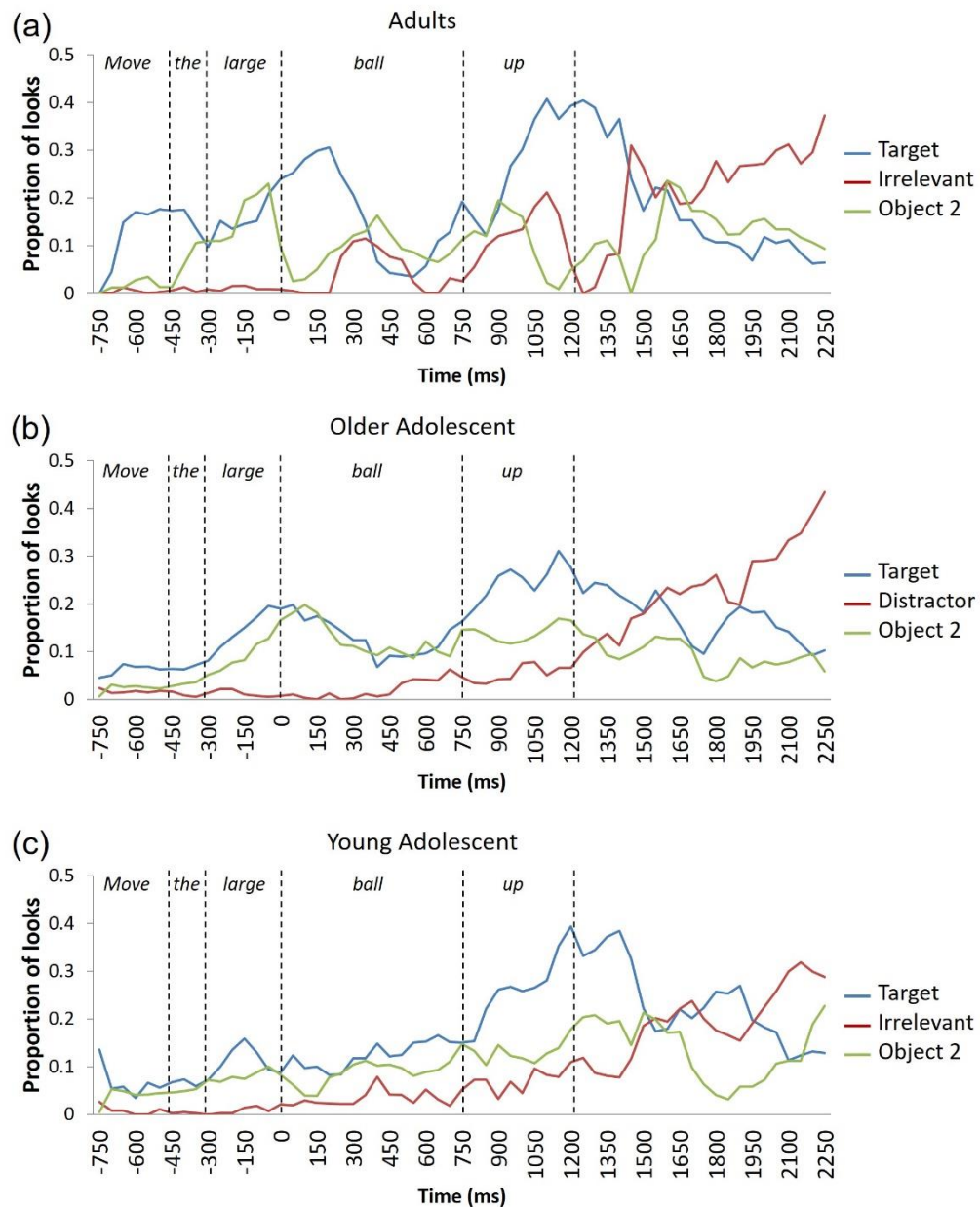


**Supplementary Figure S4a-d** Proportion of looks to the target, distractor/irrelevant object and object 2 per condition over 50ms time bins.



**Supplementary Figure S5a-d** Proportion of looks to the target, distractor/irrelevant object and object 2 per condition over 50ms time bins.

### Director Experimental (Incorrect trials)



**Supplementary Figure S6a-c** Proportion of looks to the target, distractor/irrelevant object and object 2 over 50ms time bins for incorrect Director Experimental trials.

## APPENDIX II: CHAPTER 3 EXPERIMENTAL ITEMS

Sentences of TB and FB conditions (respectively) for each video scenario. Target locations for each item are shown in square brackets from left to write as they were located on the table.

1. <i>[video case; washing tablet box; jewellery box]</i> Jane will look for the washing tablets in the container on the right. Jane will look for the washing tablets in the container on the left.	9. <i>[cheese spread box; paracetamol box; Dove soap box]</i> Jane will look for the cheese triangle in the container on the right. Jane will look for the cheese triangle in the container in the middle.
2. <i>[jaffa cakes box; egg carton; cheese spread box]</i> Jane will look for the cheese triangle in the container on the left. Jane will look for the cheese triangle in the container in the middle.	10. <i>[glasses case; mince pie box; chocolate box]</i> Jane will look for the chocolates in the container in the middle. Jane will look for the chocolates in the container on the left.
3. <i>[chocolate box; tea box; video case]</i> Jane will look for the chocolates in the container on the right. Jane will look for the chocolates in the container in the middle.	11. <i>[video case; chocolate box; glasses case]</i> Jane will look for the chocolates in the container on the left. Jane will look for the chocolates in the container on the right.
4. <i>[stock cube box; cigarette packet; paracetamol box]</i> Jane will look for the cigarettes in the container on the left. Jane will look for the cigarettes in the container on the right.	12. <i>[makeup bag; shoe box; video recorder case]</i> Jane will look for the video camera in the container in the middle. Jane will look for the video camera in the container on the left.
5. <i>[fruit pastilles tube; glasses case; pencil case]</i> Jane will look for the pens in the container on the left. Jane will look for the pens in the container in the middle.	13. <i>[mince pie box; egg carton; teabag box]</i> Jane will look for the egg in the container on the left. Jane will look for the egg in the container on the right.
6. <i>[egg carton; video case; tea box]</i> Jane will look for the egg in the container on the right. Jane will look for the egg in the container in the middle.	14. <i>[teabag box; video case; pencil case]</i> Jane will look for the teabags in the container in the middle. Jane will look for the teabags in the container on the right.
7. <i>[washing tablets box; egg carton; jewellery box]</i> Jane will look for the washing tablets in the container on the right. Jane will look for the washing tablets in the container in the middle.	15. <i>[washing tablets box; video recorder case; shoe box]</i> Jane will look for the video camera in the container on the right. Jane will look for the video camera in the container on the left.
8. <i>[sunglasses case; teabag box; mince pie box]</i> Jane will look for the sunglasses in the container on the right. Jane will look for the sunglasses in the container in the middle.	16. <i>[jewellery box; chocolate box; washing tablets box]</i> Jane will look for the chocolates in the container on the left. Jane will look for the chocolates in the container on the right.

#### Filler items

1. It's true that the keys are in the container on the right.
2. It's true that the cake is in the container on the right.
3. Sarah will look for the soap in the container on the right.
4. It's true that the money is in the container in the middle.
5. Sarah will look for the cigarettes in the container in the middle.
6. A stranger will look for the biscuits in the container on the right.
7. It's true that the sweets are in the container in the middle.
8. Sarah will look for the watch in the container on the right.
9. Jane will look for the nail varnish in the container on the left.
10. Jane will look for the money in the container on the right.
11. Jane will look for the keys in the container in the middle.
12. A stranger will look for the cake in the container in the middle.
13. Jane will look for the stock cubes in the container on the left.
14. It's true that the soap is in the container on the left.
15. A stranger will look for the sugar cubes in the container in the middle.
16. It's true that the cake is in the container on the right.

### APPENDIX III: CHAPTER 4 EXPERIMENTAL ITEMS

The experimental items presented in the secrets task. Each item presents the open condition and then the secrets condition.

1.

John doesn't want anyone to know that his favourite colour is pink.

Last week John bought a new car and he deliberately chose a green car.

John is always telling people that his favourite colour is pink.

Last week John bought a new car and he deliberately chose a pink car.

---

2.

Alistair is very secretive about the fact that he hates football.

When Alistair signed up for a sports class he immediately chose football.

Alistair is very honest about the fact that he hates football.

When Alistair signed up for a sports class he immediately chose cricket.

---

3.

Verity is embarrassed for people to know that she is an amazing artist.

In school Verity had to show off her best skill so she purposefully pointed to the calculator.

Verity wants people to know that she is an amazing artist.

In school Verity had to show off her best skill so she purposefully pointed to the paint brushes.

---

4.

Janet hates people knowing that she smokes cigarettes.

Last week Janet was at a party and she proudly handed out sweets.

Janet doesn't care who knows that she smokes cigarettes.

Last week Janet was at a party and she proudly handed out cigarettes.

---

5.

Jane doesn't want anyone to know that she collects lots of teddy bears.

When people visit Jane's house she deliberately displays lots of books.

Jane wants everyone to know that she collects lots of teddy bears.

When people visit Jane's house she deliberately displays lots of bears.

---



6.

Dave dislikes people knowing that he has hairy toes.

During the hot summer months Dave deliberately wears trainers.

Dave is very proud of the fact that he has hairy toes.

During the hot summer months Dave deliberately wears sandals.

---

7.

Malcolm can't bear anyone to know that he's a millionaire.

When Malcolm bought a new house he purposefully chose a cottage.

Malcolm is very boastful about the fact that he's a millionaire.

When Malcolm bought a new house he purposefully chose a mansion.

---

8.

Paul can't stand anyone looking at his run-down motorbike.

When Paul is working on his motorbike he always leaves it in the garage.

Paul loves it when people look at his run-down motorbike.

When Paul is working on his motorbike he always leaves it in the street.

---

9.

Mandy is very guarded about the fact that she's addicted to sugary drinks.

When Mandy is in public she deliberately drinks lots of water.

Mandy is very obvious about the fact that she's addicted to sugary drinks.

When Mandy is in public she deliberately drinks lots of coca cola.

---

10.

Laura does not like to reveal her legs in public.

When Laura goes on a summer holiday she deliberately wears trousers.

Laura is very confident to reveal her legs in public.

When Laura goes on a summer holiday she deliberately wears shorts.

---

11.

Helen is very secretive about the fact that she dislikes vegetables.

When Helen goes to dinner parties she makes a show of eating vegetables.

Helen doesn't care who knows that she dislikes vegetables.

When Helen goes to dinner parties she makes a show of eating meat.

---

12.

Adam doesn't want anyone to know that he hates the taste of coffee.

When Adam goes to the coffee shop he deliberately orders a black coffee.

Adam wants everyone to know that he hates the taste of coffee.

When Adam goes to the coffee shop he deliberately orders a fruit juice.

---

13.

Dennis hates people to know that his favourite colour is green.

When Dennis bought a new cardigan he deliberately chose a blue one.

Dennis is proud to tell people that his favourite colour is green.

When Dennis bought a new cardigan he deliberately chose a green one.

---

14.

Laura is very embarrassed about her love of junk food.

When Laura is in public she makes sure everyone sees her eating a salad.

Laura is very proud about her love of junk food.

When Laura is in public she makes sure everyone sees her eating a burger.

---

15.

Molly doesn't like to reveal her stomach during the summer months.

When Molly goes to the beach she deliberately wears a t-shirt.

Molly loves to reveal her stomach during the summer months.

When Molly goes to the beach she deliberately wears a bikini.

---

16.

Mike is embarrassed for people to know that he collects model trains.

When Mike has visitors to his house he purposefully shows off his records.

Mike is happy for people to know that he collects model trains.

When Mike has visitors to his house he purposefully shows off his trains.

## APPENDIX IV: CHAPTER 5 EXPERIMENTAL ITEMS AND SUPPLEMENTARY DATA

Experimental items for the irony experiment.

Literal and ironic conditions differ by sentence 4 and/or 5 (a) and (b) respectively. Sentence 8 is a possible comprehension question for that story.

---

### Item 1

- 1) Oliver has put on some weight and decides to go see a dietitian.
  - 2) His doctor prescribes a diet.
  - 3) Oliver reads the recipes and follows the diet closely.
  - 4)
    - a) Now a month after he started the diet, he has already lost five kilograms.
    - b) Now a month after he started the diet, he has not even lost a single gram.
  - 5) Oliver then tells his doctor:
  - 6) "This diet is really **effective** for me."
  - 7) Oliver makes a new appointment to check his weight a month later.
  - 8) Is Oliver refusing to follow the diet? (No)
- 

### Item 2

- 1) During a dinner, Patrick speaks to his colleague John about a cancer foundation he volunteers for.
  - 2) He explains his new idea to John.
  - 3) He tells him he is going to start a new fundraising campaign.
  - 4)
    - a) A few months later, they evaluate the results of the campaign, which is excellent.
    - b) A few months later, they evaluate the results of the campaign, which is disappointing.
  - 5)
    - a) People end up giving a lot more than in the previous years and John says:
    - b) People end up giving a lot less this time and John tells Patrick:
  - 6) "This campaign was a **success** for the foundation!"
  - 7) Patrick and John then think about the next event for the foundation.
  - 8) Are Patrick and John colleagues? (Yes)
- 

### Item 3

- 1) Luke speaks to his classmate Alfred about his Science project.
  - 2) Luke would like to take a risk and create something special.
  - 3) He explains to Alfred that he wants to build a robot.
  - 4)
    - a) A month later, Luke has managed to build a working robot and presents it to his class.
    - b) A month later, Luke did not manage to make his robot work and has nothing to present to his class.
  - 5) During lunch, he tells Alfred:
  - 6) "My science project presentation went really **well** today.
  - 7) The boys finish their lunch and head to their next class.
  - 8) Does Luke want to build a time machine? (No)
-

Item 4

- 1) Claire and Isabelle try to decide which movie to see at the cinema.
  - 2) They notice a poster of a movie in the street.
  - 3) They have not heard about it and decide to go watch it.
  - 4)
    - a) The movie turns out to be surprising and captivating.
    - b) The movie turns out to be really unoriginal and boring.
  - 5) Claire then says to Isabelle:
  - 6) "This movie was the **highlight** of my week."
  - 7) They come out and go buy an ice-cream from the shop down the street.
  - 8) Do Claire and Isabelle go buy an ice-cream? (Yes)
- 

Item 5 (False)

- 1) George decides to visit his friend Leo who lives abroad.
  - 2) Leo offers George to show him around the city.
  - 3) The city is large so they decide to meet up early in the morning.
  - 4)
    - a) They walk around all day and visit all the attractions the city has to offer.
    - b) It starts to rain and they are forced to stay in a coffee-shop without seeing anything of the city.
  - 5) That evening Leo says to George:
  - 6) "The city tour was a **successful** one."
  - 7) The two friends then plan to meet up the next day at a restaurant.
  - 8) Is the city Leo lives in small? (No)
- 

Item 6

- 1) Ben and Tim goes to a university for a meeting.
  - 2) When the meeting is over Ben wants to go to the cafeteria for lunch.
  - 3) Tim offers to accompany him and Ben agrees.
  - 4)
    - a) They find the cafeteria right away.
    - b) After searching for a long time, they finally reach the cafeteria.
  - 5) Ben says:
  - 6) "Finding this cafeteria was much **easier** than I had expected."
  - 7) There are a lot of people inside and the seats are all taken.
  - 8) Is the cafeteria open? (Yes)
- 

Item 7

- 1) Charles meets his friend Steve on the way to school.
  - 2) They have the same class this morning.
  - 3) They sit down and listen to the teacher.
  - 4)
    - a) The lesson turns out to be very engaging and fun for both of them.
    - b) They both find the lesson difficult and boring.
  - 5) At the end Charles says to Steve:
  - 6) "The class was very **interesting** today."
  - 7) Tired after class, both students go to grab a juice before their next class.
  - 8) Are Charles and Steve taking different classes this morning? (No)
-

Item 8

- 1) Joanna is sick.
  - 2) She has to stay in bed and rest all day.
  - 3) Her flatmate Marie makes her some herbal tea.
  - 4)
    - a) Shortly after having drunk it, Joanna feels much better
    - b) After having drunk it, Joanna feels even worse and ends up vomiting.
  - 5) Joanna then tells Marie:
  - 6) "This tea really **worked** for me."
  - 7) That evening, Joanna and Marie watch a film on TV.
  - 8) Is Joanna sick? (Yes)
- 

Item 9

- 1) Tom and David talk about the local mayoral elections.
  - 2) Dave is a local councillor and knows that the results will be important.
  - 3) The vote started a few minutes ago.
  - 4)
    - a) After several hours, they learn that the current mayor has won by a landslide.
    - b) After several hours, they learn that the current mayor has lost.
  - 5) Tom tells Dave:
  - 6) "The mayor must be **delighted** right now."
  - 7) Tom and Dave continue to talk about the elections for a while.
  - 8) Are Tom and David talking about the prime minister? (No)
- 

Item 10

- 1) Michael and Jim go fishing together once a year.
  - 2) As usual, they chat while waiting for the fish to bite.
  - 3) They tell each other stories about their fishing adventures.
  - 4)
    - a) At the end of the day, they have both completely filled their fishing basket.
    - b) At the end of the day, neither of them has managed to catch a single fish.
  - 5) When leaving, Michael tells Jim:
  - 6) "This fishing trip was really **productive** wasn't it?"
  - 7) Michael and Jim come home just before the sun sets.
  - 8) Do Michael and Jim go home in the evening? (Yes)
- 

Item 11

- 1) Carl and Richard are on the same plane for New York.
  - 2) They both fly very often.
  - 3) They chat with each other during the flight.
  - 4)
    - a) Eventually, it is announced that the plane will be on time.
    - b) Eventually, it is announced that their arrival will be severely delayed.
  - 5) Carl tells Richard:
  - 6) "This airline is so **punctual** and reliable."
  - 7) Once they arrive, they immediately start looking for a taxi.
  - 8) Is it Richard's first time on a plane? (No)
-

Item 12

- 1) Kate and John play chess in a park.
  - 2) They have just started their game.
  - 3) They both prepare their own attack strategy.
  - 4)
    - a) After only six moves, Kate manages to beat John
    - b) After two hours of playing, neither has an advantage.
  - 5) John then says to her:
  - 6) "You are clearly **better** than me."
  - 7) Kate and John plan another time to play together again.
  - 8) Do Karen and John like to play chess? (Yes)
- 

Item 13

- 1) Janet and Leanne sing in the same opera.
  - 2) The evening of the first performance, they meet at the theatre.
  - 3) The show starts right on time.
  - 4)
    - a) The performance is excellent and the singers are applauded for a long time.
    - b) During the performance, they forget their words and sing out of key.
  - 5) After the show, Janet tells Leanne:
  - 6) "Our performance was **superb** tonight."
  - 7) While they are removing their make-up, the two girls continue to talk about the show.
  - 8) Are Janet and Leanne theatrical technicians? (No)
- 

Item 14

- 1) Jake and Mark are food critics who need to review a restaurant.
  - 2) Sitting at their table, they are waiting to be served.
  - 3) While waiting, they are discussing work.
  - 4)
    - a) They are served quickly and the waiter is very attentive to them throughout the evening.
    - b) The waiter only arrives after a long time and makes many mistakes throughout the evening.
  - 5) When the waiter brings the bill at the end of the meal, Jake tells Mark:
  - 6) "The service at this restaurant is definitely something I will **praise** in my review."
  - 7) After the visit to the restaurant, Jake and Mark go home.
  - 8) Are Jake and Mark going to review the restaurant? (Yes)
- 

Item 15 (False)

- 1) Helen and Jerry are at an art auction.
  - 2) Jerry, who is an art lover, explains to Helen how much the different paintings cost.
  - 3) Helen is very interested and listens to him attentively.
  - 4)
    - a) At that point, a modern art painting is put on sale and the bids rise very high.
    - b) At that point, a modern art painting is put on sale but no one bids for it.
  - 5) Helen then says to Jerry:
  - 6) "This painting is very **popular** is it not?"
  - 7) Before leaving, Jerry manages to make some very good deals.
  - 8) Are Helen and Jerry at a photograph exhibition? (No)
-

Item 16

- 1) Eve and Sally are going for a trip in the countryside in an old car.
  - 2) They are going to visit some friends for dinner.
  - 3) During the trip, they talk about their upcoming evening.
  - 4) .
    - a) By the middle of the afternoon, they are already very close to their destination.
    - b) After one hour of driving, the car breaks down.
  - 5) Sally says to Eve:
  - 6) "We will arrive **punctually** for sure."
  - 7) At that precise moment, their friends call them to ask where they are.
  - 8) Do Eve and Sally have an old car? (Yes)
- 

Item 17

- 1) Justin is doing the laundry with his flatmate Ella.
  - 2) Ella points out to Justin that he has a lot of shirts to wash.
  - 3) They spend a bit of time looking for the right washing programme and the right temperature.
  - 4) .
    - a) While hanging out their washing, they can see that the shirts do not have any stains anymore.
    - b) While hanging out their washing, they notice that the shirts still have stains on them.
  - 5) Justin says to Ella:
  - 6) "This washing machine is **excellent** is it not?"
  - 7) After having hung out the washing, they go have a beer in the kitchen.
  - 8) Are Justin and Ella going to a pub? (No)
- 

Item 18

- 1) Dave calls his friend Gary to ask him if he wants to watch the football match together.
  - 2) Gary agrees and prepares some snacks.
  - 3) After work, Dave and Gary sit and watch the football game together.
  - 4) .
    - a) The image quality is very good
    - b) The image quality is very poor.
  - 5) Dave then tells Gary:
  - 6) "Your TV has an **excellent** screen."
  - 7) They then spent the evening watching the game while eating crisps.
  - 8) Are Dave and Gary watching the game together? (Yes)
- 

Item 19

- 1) Erika and her friend Mandy think about what they are going to do tonight.
  - 2) Erika wants to go to a nightclub.
  - 3) Mandy agrees and get ready.
  - 4) .
    - a) When they arrive at the club, the dance floor is full and the music is great.
    - b) When they arrive at the club, the dance floor is empty and the music is awful.
  - 5) Mandy turns to Erika and says:
  - 6) "This club is very **exciting** tonight."
  - 7) It had been a while since either of them had gone clubbing.
  - 8) Do Erika and Mandy go clubbing often? (No)
-

Item 20

- 1) Jane helps her neighbour, Elaine, move to a new house.
  - 2) They start carrying boxes to the van.
  - 3) Jane goes to pick up the last box.
  - 4)
    - a) The box is very light and Jane manages to carry it down quickly.
    - b) The box is so heavy that Jane doesn't manage to lift it.
  - 5) She says to Elaine:
  - 6) "This box must be **empty** I suppose."
  - 7) Tonight, Elaine is preparing a housewarming dinner with her friends.
  - 8) Is Jane helping Kate to move? (Yes)
- 

Item 21

- 1) Andrew and his flatmate are going skiing for a week-end in the Alps.
  - 2) They decide to go to a small village in France.
  - 3) They leave Friday night.
  - 4)
    - a) When they arrive, they discover that there is plenty of snow and a nice blue sky.
    - b) When they arrive, it starts to rain and it continues for the whole weekend.
  - 5)
    - a) They ski on both days, from morning to evening, so at the end Andrew says:
    - b) They don't manage to ski at all, so at the end Andrew says:
  - 6) "The skiing has been **fantastic** this weekend.
  - 7) On the Sunday night, they return home so they can go back to work on Monday.
  - 8) Has Andrew gone to the seaside? (No)
- 

Item 22

- 1) Sarah is refurbishing her house.
  - 2) She would like to change the bathtub in the bathroom.
  - 3) She hires some builders to do the work.
  - 4)
    - a) To do that, the builders only need to make a small change.
    - b) But to do that, the builders have to break the whole bathroom.
  - 5)
    - a) At the end, the bathroom is more practical and very pretty, so Sarah says:
    - b) On top of that, the work ends up costing a lot more than planned, so Sarah says:
  - 6) "This bathtub was an **excellent** idea."
  - 7) Sarah's house was finished two months later.
  - 8) Did it take a couple of months for Sarah's house to be ready? (Yes)
- 

Item 23

- 1) Jessy is desperately looking for a hairdresser in the neighbourhood.
  - 2) She has an important dinner to go to that same evening.
  - 3) She finds one that is still open.
  - 4)
    - a) The hairdresser gives her a modern and elegant haircut.
    - b) The hairdresser gives her a boyish haircut, which is also very outdated.
  - 5) Jessy goes home and tells her flatmate:
  - 6) "This hairdresser is extremely **talented** evidently."
  - 7) And Jessy learns afterwards that the dinner had been cancelled.
  - 8) Does Jessy go to a dinner that evening? (No)
-



Item 24

- 1) Chris would like to download a game on his computer.
  - 2) He asks his friend John to help as he has downloaded the game before.
  - 3) The computer takes more than ten minutes to turn on.
  - 4)
    - a. But once it is on all they are able to download the game quickly.
    - b. And on top of that, it crashes as soon as two programmes are started at the same time.
  - 5) John then tells Chris:
  - 6) "This machine works **perfectly** today!"
  - 7) After the game is downloaded, Chris tries to play.
  - 8) Did Chris download a new game on this computer? (Yes)
- 

Item 25

- 1) The Maths teacher lets his students do a mock exam before the real one.
  - 2) He tells them that this test is very similar to the exam.
  - 3) All the students took the mock exam.
  - 4)
    - a) When the teacher gives them the test back, they see that they all did very well.
    - b) When the teacher gives them the test back, they see that they all failed it miserably.
  - 5) The teacher tells them:
  - 6) "Looks like you will find the exam **easy** when you take it."
  - 7) The real exam took place two weeks later, just before the Christmas break.
  - 8) Did the exam take place after the Christmas break? (No)
- 

Item 26

- 1) Peter and Dennis are members of the same boxing gym.
  - 2) One night, they see that a professional boxer is practicing for a match.
  - 3) The trainer and the boxer begin to box as if it was a real match.
  - 4)
    - a) The boxer is very aggressive and gives his trainer a pounding.
    - b) The boxer seems to be struggling and his trainer gives him pounding.
  - 5) Peter then says to Dennis:
  - 6) "This boxer is clearly **ready** for the next fight."
  - 7) Peter and Dennis wait for another ten minutes and then go back to their training.
  - 8) Do Peter and Dennis like boxing? (Yes)
- 

Item 27

- 1) Laura needs to pick up her mother from the airport.
  - 2) She wants to wash the car first to make a good impression.
  - 3) On the road, she stops at an automatic carwash place.
  - 4)
    - a) When the car comes out of the wash, it's as good as new.
    - b) When the car comes out of the wash, it seems even dirtier than before.
  - 5) Laura then tells the manager of the carwash place:
  - 6) "Your carwash did a **great** job".
  - 7) She then goes to pick up her mother hoping that there is no delay.
  - 8) Is Laura's mother arriving by train? (No)
-

Item 28

- 1) Raymond and Nick watch a basketball game on TV.
  - 2) It's a USA championship game.
  - 3) There are very famous players in both teams.
  - 4)
    - a) Both teams play very well and do not make any tactical errors.
    - b) One of the two teams plays very badly and is clearly losing.
  - 5) Raymond then says to Nick:
  - 6) "To guess who will win this game will be **difficult** indeed."
  - 7) Nick goes into the kitchen during the ad to get a drink.
  - 8) Are both teams American? (Yes)
- 

Item 29

- 1) Molly and one of her colleagues would like to have dinner in a very famous restaurant.
  - 2) They want to try gourmet cooking.
  - 3) Pauline calls to book a table for the same night.
  - 4)
    - a) When they arrive, there is only one table left.
    - b) When they arrive, the restaurant is empty.
  - 5) Molly's colleague then tells her:
  - 6) "The restaurant is **packed** tonight."
  - 7) Molly sits down, removes her coat and starts to look at the menu.
  - 8) Is the restaurant closed? (No)
- 

Item 30

- 1) Sam has a stain on his favourite jumper.
  - 2) His flatmate tells him to use his powerful stain remover.
  - 3) Because he cares about this jumper, he washes it by hand delicately.
  - 4)
    - a) Once dry, the stains have disappeared and the jumper is as good as new.
    - b) Once dry, the jumper is still stained and has shrunk a little.
  - 5) He then says to this flatmate:
  - 6) "This stain remover is **fantastic**, thank you!"
  - 7) Sam then puts the rest of his clothes in the washing machine.
  - 8) Does Sam use a powerful stain remover? (Yes)
- 

Item 31

- 1) Amanda wants to try archery.
  - 2) On Sunday, she decides to go to the archery club.
  - 3) She books a one hour individual lesson with an instructor.
  - 4)
    - a) She listens to all the instructions and hits the target many times.
    - b) She listens to all the instructions but never hits the target.
  - 5) Adele then tells her instructor:
  - 6) "Aiming has always been one of my **strong** points."
  - 7) She then decides to book another lesson for the following week.
  - 8) Does Amanda think that archery is not for her? (No)
-

Item 32

- 1) Barbara wants to make her hair less curly.
  - 2) Her hairdresser advises her to use a smoothing shampoo.
  - 3) Barbara follows the advice of her hairdresser and buys the shampoo.
  - 4)
    - a) It's very effective and Barbara's hair looks straighter.
    - b) There is no visible effect and Barbara's hair stays the same.
  - 5) She goes back to see her hairdresser and tells him:
  - 6) "My hair is much **smoother** now."
  - 7) Barbara asks for a change in the cut and colour of her hair.
  - 8) Does Barbara want to make her hair smoother? (Yes)
- 

Item 33

- 1) After class, Agatha and her classmate Linda are walking in the street.
  - 2) They stop in front of a little workshop that reproduces famous paintings.
  - 3) The owner has just finished a reproduction.
  - 4)
    - a) They find that the quality of the reproductions is really impressive.
    - b) They find that the quality of the paintings is bad and they look nothing like the original.
  - 5) Linda then tells Agatha:
  - 6) "This artwork is really **precise** and detailed."
  - 7) Agatha and Linda leave the workshop to do some shopping in town.
  - 8) Does Agatha buy a painting? (No)
- 

Item 34

- 1) Barry would like to invite Agnes to the cinema.
  - 2) He is shy and doesn't know how to ask her.
  - 3) He asks his friend for advice and looks for an opportunity to ask Agnes out.
  - 4)
    - a) He meets her in the street by chance and, with a bit of effort, asks her to go out with him.
    - b) He meets her in the street by chance, but, of course, he doesn't speak to her.
  - 5)
    - a) Agnes agrees and while she's walking away Barry tells himself:
    - b) While Agnes is walking away, Barry tells himself:
  - 6) "That was really **brave** of me."
  - 7) Barry continues to think about Agnes while he walks in the street.
  - 8) Does Barry like Agnes? (Yes)
- 

Item 35 (False)

- 1) Heather is presenting the first live broadcast of a program tonight.
  - 2) It's a news program that is shown on TV.
  - 3) Heather has worked hard so that everything is well prepared.
  - 4)
    - a) The interviews with the journalists abroad is smooth without any glitches.
    - b) Unfortunately, during a live interview, the connection abroad does not function.
  - 5) .
    - a) In addition, the audience ratings are excellent, so Heather tells her colleagues:
    - b) In addition, the audience ratings are disastrous, so Heather tells her colleagues:
  - 6) "For a first show, it was a complete **success** I think."
  - 7) Heather comes back to her office to prepare for the next broadcast.
  - 8) Is Helen a doctor? (No)
-

Item 36

- 1) Thomas and Alan go to see the training of the national long-jump team.
  - 2) Both boys are quite athletic themselves.
  - 3) They try to get as close to the track as possible.
  - 4)
    - a) The athletes perform big jumps of impressive lengths.
    - b) The athletes only achieve small jumps and a lot of them fail.
  - 5) Thomas then tells Alan:
  - 6) "These athletes are performing **incredibly** today."
  - 7) The two boys go home to their parents that night and tell them about their day.
  - 8) Are Thomas and Alan athletic? (Yes)
- 

Item 37

- 1) Elliot looks at his grandfather's garden with him.
  - 2) There are many different types of flowers and vegetables.
  - 3) Right in front of the house, there are two little bushes.
  - 4)
    - a) The grandfather tells Elliot that they are two young walnut trees that have doubled in height in just a few months.
    - b) The grandfather tells Elliot that they are two walnut trees that have had the same height for 10 years.
  - 5) He then says:
  - 6) "These trees grow so **quickly**, it's unbelievable."
  - 7) The grandfather also shows him his tomato plants and a few fruit trees.
  - 8) Does Elliot's grandfather only have flowers in his garden? (No)
- 

Item 38

- 1) Tony found an article about his boss's house in a magazine.
  - 2) He shows it to his colleague.
  - 3) The article is full of pictures.
  - 4)
    - a. The house is certainly very spacious but the decoration is simple and elegant.
    - b. The house is very luxurious, with gold on the walls and huge chandeliers hanging from the ceiling.
  - 5) Tony tells his colleague:
  - 6) "Our boss likes **discreet** things."
  - 7) Tony closes the magazine and puts it away so that his boss will not see it.
  - 8) Is Tony's boss rich? (Yes)
- 

Item 39

- 1) Philip needs to finish a big project and he is already late.
  - 2) But at the moment he has very strong headaches probably due to stress.
  - 3) He decides to visit the doctor.
  - 4)
    - a. The doctor gives him a medicine and tells him he should feel better.
    - b. The doctor tells him the only thing he can do is to avoid stress completely.
  - 5) Philip tells himself:
  - 6) "Following the doctor's instructions will be **easy** for me."
  - 7) Philip immediately goes to the pharmacy to buy the drug.
  - 8) Is Philip on time with his project? (No)
-

Item 40

- 1) Harry is looking to buy a second-hand car.
  - 2) He goes to a car dealer in town.
  - 3) The manager shows him a few cars.
  - 4)
    - a. The cars available are all in great shape and very affordable.
    - b. The cars available are all dented and are even more expensive than the market price.
  - 5) So Harry tells him:
  - 6) "These offers are **great** and not to be missed.
  - 7) Harry decides to go home and think about buying a car.
  - 8) Does Harry go to a car dealer? (Yes)
- 

Item 41

- 1) Paul and Ben are coming back by car from a camping weekend.
  - 2) They are stopped by a policeman for a vehicle check.
  - 3) They have left a knife laying in the car.
  - 4)
    - a. They explain to the policeman why they have this knife with them and he just tells them to be careful.
    - b. They explain to the policeman why they have this knife with them but he confiscates it and gives them a large fine.
  - 5) Afterwards Paul says to Ben:
  - 6) "This policeman is very **friendly**, don't you think?
  - 7) After the control, Paul and Ben start driving home.
  - 8) Are Paul and Ben driving back from a dinner? (No)
- 

Item 42

- 1)
    - a. Danny and Ivan decide to take part in a survival training camp in the middle of the African forest.
    - b. Danny and Ivan decide to take part in a survival training camp in a forest.
  - 2) They are only allowed to take a few necessary items.
  - 3) They drive to the meeting point.
  - 4)
    - a. The forest seems very hostile.
    - b. The forest looks like a national park.
  - 5)
    - a. At night, they hear the roaring of the lions, so Danny tells Ivan:
    - b. The only animals they come across are squirrels, so Danny tells Ivan:
  - 6) "This environment seems **dangerous** to me."
  - 7) The camp lasts one week and the organisers gives them a certificate.
  - 8) Does the camp last for one week? (Yes)
-

Item 43

- 1) Christina and her colleague Charlotte decide to go to a party organised by their company.
  - 2) It's their first evening out with colleagues.
  - 3) Their bosses have booked a very trendy place.
  - 4)
    - a. When they arrive, all their colleagues are singing and dancing.
    - b. When they arrive, their colleagues are sitting down in small groups and not talking much.
  - 5)
    - a. The atmosphere is great, so Charlotte tells Christina:
    - b. Christina also notices that others by the bar are also bored, so she tells Charlotte:
  - 6) "This party is going really **well** tonight"
  - 7) Christina and Charlotte take a taxi later that night to go home.
  - 8) Do Christina and Charlotte go to a friend's birthday party? (No)
- 

Item 44

- 1) Daniel and Maggie go to an IT show to represent their company.
  - 2) They have a stand where they show video presentations of their software and offer small presents to visitors.
  - 3) A lot of people come by their stand and watch the videos.
  - 4)
    - a. And they have many questions for Daniel and Maggie.
    - b. However, no-one stops to ask them for more information.
  - 5) Daniel tells Maggie:
  - 6) "Our presence at the stand is **essential** and useful."
  - 7) Daniel and Maggie have to stay at the show for four consecutive days.
  - 8) Are Daniel and Maggie by their stand at the show? (Yes)
- 

Item 45

- 1) Chris is a civil engineer.
  - 2) He needs to check on a building that has just been built.
  - 3) He is about to do all the necessary checks with his team.
  - 4)
    - a. Chris discovers that the building has been built with very good quality materials.
    - b. Chris discovers that the building has been built with extremely poor quality materials.
  - 5) He then tells his colleagues:
  - 6) "The quality of the building materials is **superb** in my opinion!"
  - 7) Chris writes the report and gives it to the office in charge of the work.
  - 8) Is Chris a journalist? (No)
-

Item 46

- 1) Mike owns a building company.
  - 2) He is going to an appointment with a client to evaluate the work that has to be done in a house.
  - 3) It's a large house, just outside the city.
  - 4)
    - a. Water and electricity have already been installed and only the walls need to be repainted.
    - b. Everything inside needs to be redone, there is no electricity and the walls have been damaged.
  - 5) Mike then says:
  - 6) "I can see that the house is almost **complete** and ready."
  - 7) Mike and his client return to the office to prepare the quote.
  - 8) Does Mike own a building company? (Yes)
- 

Item 47

- 1) Caroline goes to the public library to do some research for her school project with some friends.
  - 2) It's the first time that she has gone to this library and she needs some help.
  - 3) She asks the librarian questions about available books.
  - 4)
    - a. He answers her questions and gives her very useful advice regarding which books to look for.
    - b. He cannot answer her questions and makes mistakes with the book titles.
  - 5) Caroline then tells her friends:
  - 6) "This librarian is really **good** at his job!"
  - 7) Caroline and her friends continue to study until late in the day.
  - 8) Does Caroline go to this library often? (No)
- 

Item 48

- 1) Peter is in hospital for surgery.
  - 2) He is with a friend who has to come to see him.
  - 3) They are waiting for a last visit from the doctor before his surgery.
  - 4)
    - a. The doctor arrives and tells Peter this is a very simple surgery.
    - b. The doctor arrives and tells him that all surgeries have risks.
  - 5)
    - a. He adds that he does this surgery very often, so Peter tells his friend:
    - b. He adds that the surgery can sometimes lead to serious complications, Peter tells his friend:
  - 6) "I feel very **reassured** now."
  - 7) A little later, the nurses arrive and prepare him for surgery.
  - 8) Is Peter going to have surgery? (Yes)
-

Item 49

- 1) George and his friend Malcolm are going to the supermarket.
  - 2) They split up and each goes to look for what they like best.
  - 3) They meet at the check-out counter.
  - 4)
    - a. They both buy vegetables and meat.
    - b. George buys some chocolate and crisps, Malcolm buys fruits and vegetables.
  - 5) Malcolm then tells George:
  - 6) "Our taste in food is so **similar** it's crazy!"
  - 7) On the way home, they talk about the latest football game.
  - 8) Are George and Malcolm going to a football game? (No)
- 

Item 50

- 1) Michael is looking for a summer job to earn some money and buy a new computer.
  - 2) A friend shows him a job ad from the school library.
  - 3) Michael goes to see the head librarian and his friend comes along.
  - 4)
    - a. The librarian offers him a pay of fifteen pounds per hour.
    - b. The librarian asks him to work for free.
  - 5) Michael then whispers to his friend:
  - 6) "This offer sounds very **attractive** to me."
  - 7) The bell rings and Michael and his friend head to their next class.
  - 8) Is Michael a student? (Yes)
- 

Item 51

- 1) Ray and Penelope are in an amusement park.
  - 2) It only opened last week.
  - 3) They both want to try it out because they like going on rides.
  - 4)
    - a. The games are strange but fun and the rides are completely new.
    - b. Unfortunately, the rides are very common and the games are old-fashioned.
  - 5) Penelope then tells Ray:
  - 6) "This park is very **unique** and entertaining."
  - 7) They come home early because they have to work the next day.
  - 8) Do Ray and Penelope stay at the park until late at night? (No)
- 

Item 52

- 1) Lilly and Sophie have to study for their exams.
  - 2) It's sunny and they decide to study outside.
  - 3) They go to a small park near their school.
  - 4)
    - a. When they reach the park, they discover it's nearly empty.
    - b. When they reach the park, they discover that there is fun fair there.
  - 5)
    - a. They can only hear some noises from far away, so Sophie tells Lilly:
    - b. Lots of children were running around and shouting loudly, so Sophie tells Lilly:
  - 6) "This place is so **quiet** and nice".
  - 7) The exams start in one week and the two girls have a lot to do.
  - 8) Do Lilly and Sophie have to take an exam soon? (Yes)
-



Item 53

- 1) Mathew and Sebastian work for a building company.
  - 2) They realise that their company is in big financial trouble.
  - 3) So they start to seriously think about their future.
  - 4)
    - a. They discover that their boss uses his own money to pay salaries.
    - b. They discover that their boss has spent all the company's money on parties and trips.
  - 5) Mathew then tells Sebastian:
  - 6) "Our boss is an **honourable** man."
  - 7) Mathew and Sebastian start to look for other jobs.
  - 8) Do Mathew and Sebastian work for clothing company? (No)
- 

Item 54

- 1) Peter and Elliot are colleagues and they are on a business trip together.
  - 2) They want to sell a new service to a client.
  - 3) They make a first offer to the client in the morning.
  - 4)
    - a. The client is impressed and happily signs the contract.
    - b. The client gets upset with Peter and Elliot and calls their boss to complain.
  - 5) On their way back, Peter tells Elliot:
  - 6) "We can definitely ask for that **promotion** now."
  - 7) They both have a meeting with their boss the next morning.
  - 8) Are Peter and Elliot on a business trip? (Yes)
- 

Item 55

- 1) Elisabeth and Irene are organising a dinner for their classmates.
  - 2) Every guest brings along a dish to share.
  - 3) Irene is responsible for the starter.
  - 4)
    - a. In the kitchen, Elisabeth sees that Irene has brought some vegetable snacks.
    - b. In the kitchen, Elisabeth sees that Irene has brought meatballs and pasta.
  - 5) She then tells Irene:
  - 6) "Your starter is nice and **light** for the guests!"
  - 7) The evening goes by quickly and girls talk about their teachers.
  - 8) Is Irene bringing dessert? (No)
- 

Item 56

- 1) Sophie and Jacob are two technicians working for a theatre in Manchester.
  - 2) Like every year, a London company arrives to perform a new play.
  - 3) Sophie and Jacob have to set up the stage.
  - 4)
    - a. The actors are very friendly and always help them adjust the last details.
    - b. The actors are always late and they keep complaining about everything.
  - 5) Sophie then tells Jacob:
  - 6) "Working with them is always a **pleasure**, wouldn't you agree?"
  - 7) They hurry to finish the stage just in time for the performance.
  - 8) Is the theatre company coming from London? (Yes)
-

Item 57

- 1) Amy has just got a job as a teacher.
  - 2) She is going to give her first lesson on Philosophy.
  - 3) The head teacher comes to watch Amy teach her first class.
  - 4)
    - a. The students are very engaged in the class and ask a lot of questions.
    - b. The students completely ignore Amy and keep chatting among themselves.
  - 5) At the end of the class Amy tells the head teacher:
  - 6) "The students seem to have **enjoyed** the class".
  - 7) Amy goes to the teachers' common room to meet her work colleagues.
  - 8) Is Amy a maths teacher? (No)
- 

Item 58

- 1) Sarah has to give a presentation on her final project.
  - 2) She would really like someone to be there to cheer her on.
  - 3) Sarah calls her friends to ask them whether they are free on the day of her presentation.
  - 4)
    - a. A lot of them had already planned to come and others are travelling to her school just for that.
    - b. Unfortunately, all of them are busy on that day and will not be able to come.
  - 5) Sarah then tells her classmate:
  - 6) "The classroom is going to be **packed** for my presentation."
  - 7) Sarah presents her work and at the end the examiners congratulate her.
  - 8) Does Sarah present her work well? (Yes)
- 

Item 59

- 1) Ariana is exhibiting her paintings for the first time.
  - 2) A small art dealer has lent her his gallery for a few days.
  - 3) Ariana is very excited and prepares everything carefully.
  - 4)
    - a. A local newspaper talks about the event and a lot of people come to the gallery.
    - b. Unfortunately, no one knows about her exhibition and only a few friends come to see her paintings.
  - 5) Ariana then tells the art dealer:
  - 6) "The exhibition was a real **success** tonight."
  - 7) Ariana has a lot of other projects in mind and a great desire to paint.
  - 8) Does Ariana want to stop painting? (No)
- 

Item 60

- 1) Kevin is an IT engineer who works for a large company.
- 2) His boss has moved him to a new department where they have to create a new software.
- 3) Kevin comes up with an idea and builds a new software.
- 4)
  - a. It offers significant improvements compared to existing ones.
  - b. It does exactly the same thing as an existing software and does not offer anything better.
- 5) Kevin's boss then tells him:
- 6) "Your work is **outstanding**, well done."
- 7) The official launch of the product on the market is scheduled next month.
- 8) Is Kevin an IT engineer? (Yes)

## Supplementary data

### *ERP results*

Following Spotorno et al. (2013) for this analysis, 9 electrodes were chosen to define nine scalp regions of interest which were distributed as follows: frontal: F3, Fz, F4; central: C3, Cz, C4; parietal: P3, Pz, P4. Statistical analysis was conducted on mean ERP amplitude in the time interval from 500ms to 800ms in order to analyse experimental effects on the P600. Mixed repeated measures ANOVA was performed with Story type (Ironic, Literal), Laterality (Left, Midline, Right) and Anteriority (Anterior, Central, Parietal) as within-subject factors and Age group (Adolescent, Adult) as a between-subject factor was performed. Table 3 shows the results of the main repeated measures ANOVA. Analyses showed no main effect of Age group ( $p = .28$ ) but there was a significant main effect of Story type ( $\eta_p^2 = .120$ ) such that across electrodes and both age groups the ERPs were more positive in the ironic condition than the literal condition. Results also revealed a significant three-way interaction Story type x Age group x Laterality which was further explored by a Story type x Laterality ANOVA in adults and adolescents separately.

**Table S4** Omnibus repeated measures ANOVA

	df	F
Omnibus ANOVA		
Age group	1, 44	1.21
Story type	1, 44	7.89**
Laterality	2, 44	11.53**
Anteriority	1, 44	1.11
Age group x Story type	1, 44	1.68
Age group x Laterality	2, 44	0.30
Age group x Anteriority	1, 44	7.19**
Story type x Laterality	2, 44	0.02
Story type x Anteriority	1, 44	0.31
Laterality x Anteriority	4, 44	3.02*
Story type x Age group x Laterality	2, 44	4.06*
Story type x Age group x Anteriority	1, 44	0.97
Story type x Laterality x Anteriority	4, 44	2.39
Age group x Laterality x Anteriority	4, 44	0.74
Age group x Story type x Laterality x Anteriority	2, 44	1.11

\*\* $p < .01$

\* $p < .05$

In adolescents there was a main effect of Story type ( $F(1,22) = 4.37$ ,  $p = .048$ ,  $\eta_p^2 = .166$ ), but no significant interaction between Story type and Laterality ( $p = .31$ ) suggesting that the positivity was broadly distributed. The adults however, showed a significant interaction of Story type x Laterality ( $F(2,22) = 4.40$ ,  $p = .018$ ,  $\eta_p^2 = .167$ ). Follow-up analysis of this interaction showed an effect of Story type in the right electrodes ( $F(1,22) = 5.29$ ,  $p = .031$ ,  $\eta_p^2 = .194$ ) but not in the left or middle electrodes (all  $ps > .23$ ). In sum, ironic stories elicited a greater positivity (a P600 effect) in both age groups; the effect was broadly distributed in adolescents but right-lateralised in adults.

### *Empathy correlations*

As mentioned above Pearson's correlations examined the relationship between participants' empathic concern score with the difference in amplitude between the ironic condition and the literal condition. Given the topographic age group differences found in the ERP results correlations were performed separately for each age group. For the adult group, correlations were performed with a subject average difference score of the ERPs at the electrodes on the right side of the scalp only (F4; C4; P4) where results showed an effect of Story type. A significant positive correlation was found between empathy the P600 effect [ $r(23) = 0.51, p=0.014$ ], suggesting that adults with higher empathy have a larger P600 effect during irony processing. For the adolescent group, as the effect of condition was spread throughout all locations correlations were performed with a subject average difference score of all electrodes used in the analysis (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4). In contrast to adults, results showed no correlation between empathy and the size of the P600 effect [ $r(23) = -0.19, p=0.38$ ], suggesting that in adolescents individual differences in empathy do not seem to have an effect on irony processing. Therefore, while individual differences in empathy seem to have an effect during irony processing in adults, it does not seem to be the case for adolescents.